VOL. XXIV, No. 4, March—April, 1951

MATHEMATICS magazine

MATHEMATICS MAGAZINE

Formerly National Mathematics Magazine, founded by S. T. Sanders.

Address editorial correspondence to Glenn James, special papers to the editors of the departments for which they are intended, and general papers to some member of the Executive Committee.

Manuscripts should be typed on 8½" x 11" paper, double-spaced with 1" margins. We prefer that, in technical papers, the usual introduction be preceded by a *foreword* which states in simple terms what the paper is about.

The Mathematics Magazine is published at Pacoima, California by the managing editor, bi-monthly except July-August. Ordinary subscriptions are $3.00, sponsoring subscriptions $10.00, single copies 65¢. Reprints: 25 free to authors; 100 @ $1.25 per page with a minimum of $6.00. All reprints are bound.

Subscriptions and related correspondence should be sent to Inez James, 14068 Van Nuys Blvd., Pacoima, Calif.

Advertisers should contact the managing editor.

Entered as second-class matter March 23, 1948 at the Post Office, Pacoima, California under act of congress of March 8, 1876.

SPONSORING SUBSCRIBERS

CONTENTS

OPTIMUM SELECTION AND OPTIMUM REGISTRATION

(EQUILIBRIUM OF AWARENESS)

Andrew Pikler

I.

The Optimum Selection Problem

In the traditional theory of Marginal Utility in Econometrics we have to do with a *limited stock* of commodity or a *limited amount of income* and owing to this situation of shortage we are supposed to materialize *optimum allocation*. In the present study we shall discuss a somewhat different hypothesis. It will be assumed that there exists an abundance of various types of available commodities, however, we dispose of only a limited volume (f.i. a limited space of storage) for locating these commodities. The task now consists of selecting the most valuable units of commodities with which we are supposed to fill out the units of the available volume.

We shall discuss the *optimum selection* problem as a mathematical model which has manifold applications outside the fields of Economics. In *social problems* we often meet the task of filling out a "contingent" with the best units. Probably the *Theory of Games* can also be formulated in terms of optimum selection. Within a closed section of a game, retrospectively, the player was supposed to have selected the best units of *possible* moves and to have allocated them within the various "fields" of the game. A similar concept of the games necessitates the formulation of the *Utility* of the moves and of the *limitations* of the moves in terms of the rules of the game.

Optimum selection is obviously operative also in *Physiology*. A living organism is constantly facing the task to materialize Optimum Selection in utilizing the ingredients of the consumed food with regard to the biochemical limitations. In the structure of huge *animal societies* Optimum Selection materializes through the struggle for life in terms of the "survival of the fittest".

Optimum Selection has a very typical application in *psychophysics*, *i.e.* in the Psychology of the stimuli and the sensations. We shall discuss the mathematical implications of Optimum Selection from an abstract model with the didactical purpose to further the understanding of the psychophysical phenomena.

II.

Mathematical Data

A.) *Initial data.*

1.) q_I, q_{II}, ... q_N are the quantities of the different types of

commodities. These quantities are measured in terms of their specific units (weight, number of pieces, etc.)

2.) V is the available total volume. In more general terms, it is the "restricted contingent" which is supposed to be filled out on the basis of utility considerations. It is measured in the specific units of volume, *e.g.* cubic feet, if a problem of storage or of transportation is concerned.

3.) $f_1(q_1)$, $f_2(q_2)$, ... $f_n(q_n)$ are Utility Functions for the different commodities. The easiest approach to the Optimum Selection Theory is to presume for the time being that the utility of a certain quantity of commodity for a person in a given moment and for a given period, is measurable in terms of "utils". (Irving Fisher.) "Util" is the imaginary unit for measuring utility. It is generally assumed that Utility is a logarithmic function of the quantity of commodity possessed, *i.e.* utility is increasing at a slower rate than the stock of the commodity. —For the deeper concept of measurability see Note II — and Figure I.

4.) $\phi_1(q_1)$, $\phi_2(q_2)$ ..... $\phi_n(q_n)$ are the functions for "expansion." They give the functional relationship between quantity of commodity and the volume occupied by it. In the simplest case, expansion is linear; the more commodity the more volume. In hypotheses connected with storage and transportation, the exploitation of space improves with increased quantities; expansion is materializing, therefore, on a more favorable, namely the logarithmic, scale.

B) *Data derived from the initial data.*

5.) The volumes occupied actually by the available stocks can be calculated from the data under 1) and 4).

6.) Utility can be attached not only to the quantities of commodities, but also to the "utilized volumes," namely in terms of each type of commodity. This is an implicit result, inasmuch as utility is a function of the quantity of commodity, as is also the occupied volume itself. If we are assuming the linearity of expansion (see 4), the relationship between utility and utilized volume is very simple. The utility (U) of the occupied volume in terms of a chosen commodity is also translated in a logarithmic relationship,

$$U = \ln v$$

The utility of the occupied volume is measured by the *logarithm* of the volume itself.

III.

Unknowns of the Problem.

In the Optimum Selection problem we are asking ultimately for the final division of the Volume pertaining to its different uses. Since we have "n" types of commodities, Total Volume V will be divided into "n" parts

$$V = v_I + v_{II} + \dots v_N$$

Our equation points out N unknowns.

All the other unknowns of the problem have a merely *derived* character:

1) If we can obtain the solutions for the occupied volumes v_I, v_{II}, ... v_N, we shall be able to calculate also the respective quantities of the selected commodities, namely with the aid of our "functions of expansion." (See 4) on p. 176). For the brevity of terminology, let us call these quantities which are actually included into the selected units the "*included quantities*."

2) Also the quantities not included ("excluded") into the accomplished selection can be easily calculated since

$$q = q_{incl} + q_{excl}.$$

IV.

The Conditions of Optimum Selection.

Since we have N unknowns and only *one* equation, we need $(N - 1)$ additional equations. This set of $(N - 1)$ equations is furnished through the following considerations:

There exists a function U which is supposed to be maximized in the situation of Optimum Selection. It is the Joint Utility Function for the Occupied Volumes, or briefly the Joint Utility Function. The utilization of the volume is supposed to materialize the maximum of total Utility:

$$U_I + U_{II} + \dots U_N = U \quad max.$$

Maximization is reached if all the first derivatives with respect to the occupied volumes are equal:

$$\frac{dU_I}{dv_I} = \frac{dU_{II}}{dv_{II}} \dots = \frac{dU_N}{dv_N}.$$

This condition points out $(N\text{-}1)$ additional equations.

Strictly speaking, we also have to show that the extreme value is a maximum and *not* a minimum. In view of the shape of the operative functions (continuous increase of Utility on the logarithmic scale), this implication is obvious.

V.

Discussion of the Optimum Situation.

1.) The "marginal utilities with respect to the occupied volumes" over all the occupied volumes are equal. Optimum selection means *marginal*

utilization of the occupied volumes *on equal level*. A similar formulation of the optimum conditions is immediately evident by virtue of the principle of the "virtual displacements". It is *not worthwhile* for a person to "move" from a similar situation. (Marginal Law)

2.) Let us introduce the idea of the "*Average Utility*" of the volume unit within a *layer* of the total volume, alloted to one type of commodity.

$$\frac{\int_0^v U\,dv}{v}$$

As a matter of fact, the *average utilities* over the different layers alloted to the different commodities are *not* equal.

3.) Let us construct now the following concept: "average utility of the volume unit over the total volume".

$$\frac{\text{Grand Total of Utilities}}{\text{Total Volume}} = \frac{\int_0^{v_I} U\,dv + \int_0^{v_{II}} U\,dv + \int_0^{v_{III}} U\,dv + \ldots\ldots}{V}$$

This concept is translating the "average utilization" in view of the total volume. It is obvious that in the optimum situation the general average of the utilization is brought to its *maximum*.

4.) In surveying the situation from the point of view of the marginal utilities of the commodities themselves, it is understood that the *marginal utilities* of the included commodities are *not equal* in general. Marginal utilities are *different* for the different commodities.

5.) In all our previous hypotheses the utility of a commodity was considered as a function of only one independent variable, *i.e.* the quantity of the commodity. In numerous hypotheses connected with the idea of *complementarity* (example: bread and butter) and *substitution* (example: butter and fat), utility is a function of two or *several* variables:

$$U = f(q_I, q_{II}, q_{III} \ldots)$$

Similar implacations will accordingly modify the conditions for optimum selection. (See Figure I)

SECOND PART.

I.

The Optimum Registration Problem in Psychophysics.

An optimum problem of quite specific character appears in numerous fields where an intelligent *completeness* of the information is required

in a *restricted* framework. Which type of information shall we select and how much from each type? Similar problems can be classified as problems pertaining to Optimum Selection, whereas optimum selection is sought with the objective to materialize *Optimum Registration*. Some conception concerning the *valuation* of the selected information, *i.e.* its *importance*, will be indispensable. No optimum selection problem can be solved without formulating the operative concept of Utility.

Psychophysics is the field in which Nature itself is supposed to work out an Optimum Selection of the Stimuli in order to present their Optimum Registration in our awareness. We can assume that Nature wishes us to have a *valuated* picture concerning the outside world. The psychophysical mechanism of our sensations takes care of an information service not merely on qualitative lines: it secures not only the identification of the operative stimuli (*e.g.* light and sound) with certain sensations (vision and hearing.) All the operative stimuli and the resulting sensations are *quantitative* issues also, both of them have their definite *intensities*. It is reasonable to assume that Nature constantly carries out a certain optimum seleotion of the operative stimuli in view of the fact that the *scope of our sensory awareness is restricted*. Let us try to formulate the basic problem of Psychophysics, namely the *composition and the equilibrium of the sensory awareness* in terms of a mathematical problem of *maximization*.

Let us start from the assumption that in a given moment a person is facing the following set of operative stimuli in the different sensory fields (light, sound etc.).

$$S_I, \quad S_{II}, \; \ldots\ldots \; S_N$$

If there exists a *Joint Utility Function* (see p. 177) for the selected sets of stimuli, this *joint characteristic function* will have the following structure

$$U = F(s_I, \; s_{II} \; \ldots \; s_N) \; \ldots\ldots 1,$$

This characteristic function is supposed to be maximized in a situation which might be considered as the *equilibrium of awareness*. This maximization must be carried out with regard to a restriction which we might formulate as the *limitation of our awareness*. Our *awareness* cannot register the total of the operative stimuli, namely

$$S_I, \; S_{II} \; \ldots \; S_N$$

Only a selected set of stimuli can enter into our awareness, *i.e.*

$$s_I, \; s_{II} \; \ldots \; s_N$$

Each member of the selected set is smaller than the operative amount of

the stimulus itself.

$$s_I < S_I, \quad s_{II} < S_{II} \ldots s_N < S_N$$

Can we specify somehow what we consider the *limitation of our awareness?*

Let us anticipate from the traditional psychophysical theory the following consideration: For each type of stimulus there exists a corresponding function of "sensation—intensity" , namely

$$f_I(s_I), \quad f_{II}(s_{II}), \ldots f_N(s_N)$$

The total of the registered sensation—intensities is a *constant*. This constant represents the "total scope of our awareness" .

$$f_I(s_I) + f_{II}(s_{II}) + \ldots f_N(s_N) = \text{Const} \ldots\ldots 2,$$

Can the Joint Utility Function and the Limitation of the Awareness be mathematically specified in a more tangible way? Are there experimentally accessible implications in Psychophysics which justify the *introduction of functional operations into Psychophysics?* We shall undertake such an attempt. The generally accepted facts incorporated in the fundamental *psychophysical laws of Weber and Fechner* will serve as a starting point. However, it is indispensible to anticipate some methodological considerations.

II.

Methodological Considerations.

1.) All our measurements, physical or psychical, are within the range of the *elementary quanta*. Measurement in Psychophysics, therefore, means simply the *counting of the elementary quanta.*

2.) The intensity of a sensation is measured in terms of the "J.N. D." -s, *i.e.* the number of the *just noticeable differences* which are supposed simply to be *counted*.

3.) All our functions are *quanta functions*. Their variables are varying in terms of discontinuous increments, namely the elementary quanta.

4.) Our Joint Utility Function will have the character of a so-called *Index Function*. For the explanation and illustration of this important assumption let us quote an analogy with Utility (Ophelimity) in Econometrics. In formulating our Joint Utility Function we should not expect that there exists just one *unique*, concrete function which will satisfy the underlying requirements. We should *not* attempt to establish one unique scale for measurable utility with numerically determined values. Also we shall not postulate that the maximum problem should be numerically solvable for one unique value of the maximum utility itself.

Instead of similar erroneous presumptions which prove correct only

in physical measurements on the macroscopic scale, we are adopting the following consideration:

There exists an *entire family of utility functions* which is supposed to hold for the underlying requirements. This family of functions, however, must have a common feature: each member of the family shall attain its maximum at the same point with regard to a given area of the variations, namely at the point

$$P(s_I, \quad s_{II} \dots s_N)$$

5.) Can the *scope of our sensory awareness* be handled as a quantitative entity also? In principle our answer is an affirmative one. For example visual brightness shows 572 steps for illumination from absolute darkness to the maximum brilliance that the eye can stand. Furthermore, there are about 366 discriminable different intensities for a tone of 2048 cycles per second. In principle, we can assume that the scope of the sensorial awareness can be *measured* by *counting* the J.N.D.-es (the just noticeable differences) with which the sensorial awareness is completely filled out within one given field of stimulus sensation.

6.) We have to survey and to formulate, the *psychophysical functions* from which the mathematical issues of our maximization problem are supposed to be derived. If we have accepted the idea of quanta-functions, let us also not refrain from carrying out *operations* also concerning these functions. Differentiation, integration, calculation of the mean value in terms of the Integral Calculus, Maximization etc. are justified mathematical operations.

III.

Survey of the Psychophysical Functions

1.) *First function*. "Differential sensitivity of the first type." Stimulus increases, the sensation increment ("differential sensation") decreases. This relationship is covered by Weber's fundamental psychophysical law which is translated in our statement in terms of a continuous function. It is assumed that each J.N.D. ("just noticeable increment" of sensation), each ΔI is equal. We thus arrive to a formulation of the law which translates the idea of the "diminishing returns" of the stimuli. It is the law of the "decreasing rate of differential sensitivity over a certain range of stimulus.

$$r = \frac{\Delta I}{\Delta s} = \frac{a}{s}$$

r = rate of differential sensitivity

ΔI = sensation increment

Δs = stimulus increment

a = constant

The above function is illustrated by a Hyperbola (See Figure II). It is a function accessible for experience and for further derivations. We only need to test the numerical value of the so-called *Weber fraction* for a certain person at a certain moment over the different ranges of the different stimuli. (See Note III).

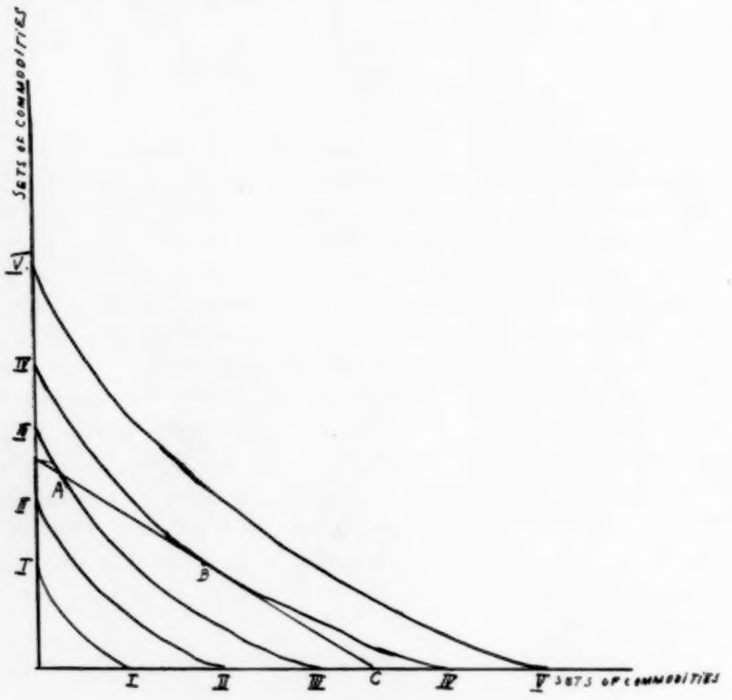


Figure I
An "indifference map".

The indicators I II III IV V represent "levels of utility". Point *A* represents an arbitrarily selected set of commodities. The line $\overline{ABC}$ is the "selection line". It is the locus of sets of commodities compatible with the volume restriction.

Point *B* represents the optimum selection of commodities. The selection line *ABC* is a tangent to indifference curve IV in point *B*.

It will be opportune to introduce a new concept. If our awareness is registering (selecting) the stimulus S, after accomplishing selection, it will be of interest to ask for the numerical value of the *last ordinate*. The last ordinate of the curve expresses the "*marginal rate of sensitivity*", or quite briefly the "*marginal rate*". *Marginal rate is the differential sensitivity for the last stimulus increment admitted to our awareness.* (See Figure II)

2.) *Second function. Differential sensitivity, second type.* Detrimental stimuli (work, energy loss, attacks against the body or the

organism) operate in the reverse way. Additional stimuli are translated in successively increasing increments of our sensation. The curve of the differential sensitivity is continuously ascending and is illustrated by a Power Curve. The function itself is a *power function.*

$$\frac{\Delta I}{\Delta s} = as^n$$

Once again we can accordingly formulate our concept for the "marginal rate of sensitivity" too.

3.) *Third function. Total sensation, first type.* Following *Fechner's* considerations, we undertake integration. If we integrate the first function for differential sensitivity, we arrive at the function of the *total intensity of the sensation.*

$$\int \frac{\Delta I}{\Delta s}\, ds = I = \int \frac{a}{s}\, ds = a \ln s + \text{const.}$$

The total intensity of sensation is a logarithmic function of the stimulus. It can be easily shown that the Integration Constant covers the idea of the absolute threshold value of the stimulus. (See Note IV.)

4.) *Fourth Function. Total sensation of the second type.* The total sensation pertaining to a finite range of the detrimental stimuli (see second function) can be established once again through integration. If Function II. for differential sensitivity is integrated we arrive at a Power Curve which is *steeper.*

$$I = \int a\ s^n ds = a\,\frac{s^{n+1}}{n+1} + \text{Const.}$$

Our Integration Constant will once again cover the idea of the absolute threshold for the detrimental stimulus.

IV.

The Joint Utility Function and its Maximization.

From our first two functions for differential sensitivity (p. 182–183) we wish to derivate deductively the function for Joint Utility. (p. 177). Perhaps in some future stage of the physiological science we shall be able to measure this Utility through objective data. This Utility is supposed to cover the following ideas: "biological importance of becoming aware of a certain stimulus". It is the *Utility of Information.* At present we are referred to our subjective data, namely to our sensations themselves. They reflect in a quite noticeable way the issue which we called the Utility of the Information, or the *biological valuation of a stimulus.* In the Chapter of Psychology concerning differential sensitivity it is a very generally adopted consideration to state that the biological importance of a stimulus increment is reflected by the respective "rate of sensitivity". Nature makes differential sensitivity

keen over a certain stimulus increment because the information about this stimulus increment is of a big *interest* for us. Differential sensitivity is small if the importance of the information concerning a stimulus increment is *small.*

How can we compute Total Utility over a certain finite range? Our suggested answer is the following.

It is the *average of the ordinates* (see functions 1 and 2, p. 182-183, see also Figure II and III) over a finite range which is decisive for the total importance of the range itself. This average can be calculated with the aid of the Mean Value Theorem of the Integral Calculus.

$$U = \frac{\int_0^s r\,ds}{s} = \frac{\int_0^s \frac{dI}{ds}\,ds}{s}$$

The result is illustrated on Figure II and III.

If a similar consideration holds for one type of stimulus over a certain range, we now can formulate the Joint Utility Function for all selected stimuli over the respective ranges. By virtue of the principle of *additivity of utilities* we can say: The Joint Utility of all stimuli equals the total of the single utilities.

$$U_{\text{total}} = \frac{\int_0^{s_I} r_I\,ds}{s_I} + \frac{\int_0^{s_{II}} r_{II}\,ds}{s_{II}} + \dots \frac{\int_0^{s_N} r_N\,ds}{s_N} \ \max$$

This is the function which is supposed to be maximized under the restriction already established, namely

$$f_I(s_I) + f_{II}(s_{II}) + \dots f_N(s_N) = \text{Const.}$$

Already we are able to specify the nature of this restriction. Namely the individual operative functions $f_I(s_I)$, $f_{II}(s_{II})$ ... $f_N(s_N)$ can be classified either as pertaining to the third type of the psychophysical functions (Logarithm curves of Fechner) or to the fourth type of the psychophysical functions. (Power curves).

Can a general psychophysical theorem be established with regard to the idea of optimum selection?

It would be tempting to invoke simply the analogy with the theorems of optimum selection as exposed in the first part of this study. Commodity corresponds to stimulus, selected commodity to registered stimulus, total volume to total scope of awareness, the hyperbola of the marginal utility curve to the first psychophysical function etc. However, there are difinite discrepencies in the mathematical implications of the two model-constructions. The solution of the optimum problem must be carried out on independent lines. (See Note V)

Let us restrict ourselves to the *general* aspects of the psychophysical information theory with regard to maximization. There is an obvious physical analogy on hand, namely the analogy with *Entropy* in thermodynamics. In both of the hypotheses involved, in Psychophysics and in Thermodynamics, we meet a *characteristic function which is supposed to be maximized just in the final situation of a composite system.* This final situation is considered as that of the *equilibrium*. Obviously, we can introduce also into Psychophysics the concept of equilibrium, namely that of the "*equilibrium of awareness*". Under the effect of a constant stimulation from the outside world, our awareness has the tendency to materialize an *equilibrium*.

Furthermore, from our mathematical assumptions it is understood that *the equilibrium situation is covered by a system of equations.* There are "*n*" simultaneously operative stimuli; there are "*n*" corresponding sensation-intensities. The number of the equations furnished by the problem of maximization must be *equal* to those of the unknowns, otherwise there would be an *indeterminacy*.

Generalization of the Equilibrium Theory.

If two or several stimuli act simultaneously, it was assumed in all our previous hypotheses that the respective intensities of sensations are simply added in our awareness. This is the *principle of additivity*. However, in numerous hypotheses it is clear that the principle of additivity does *not* hold. The following hypotheses for non-additivity must be covered mathematically:

a.) *Mutual neutralization*. (*E.g.* painful stimulus and pain reliever.)

b.) *Masking* of one stimulus by the other. (The higher musical tone suppresses the sensation of the lower. "Shifting of the threshold.")

c.) *Mutual augmentation*. Two pleasure sensations meet. The resultant "pleasure" is even more intensive than the pleasure intensities attributed separately to the two stimuli. (Example: consumption of complementary goods.)

d.) *Tertiary effects*. Two simultaneously associated stimuli can create a third sensation besides the other two. (Example: two simultaneous musical sounds produce the awareness of the interval.)

In the mathematical treatment of simultaneously interacting stimuli we must introduce functions with multivariables. The resultant intensity of sensation is the function of a set of interacting stimuli

$$I = f_I(s_I, s_{II} \dots s_N)$$

This new implication in the model will naturally be translated in all the issues connected with maximization.

VI.

Dynamical aspects of the equilibrium theory

1.) *Attention.* If our attention is focused on one particular stimulus, the operative psychophysical function itself is modified. Through attention we are determining *spontaneously* the shape of the operative curve. Differential sensitivity becomes keener. The equilibrium of our awareness will be adequately shifted. Furthermore special cases will arise if attention is divided among different stimuli.

2.) Other typical cases for shifting of the equilibrium are:

a.) *A change in the stimuli.* The equilibrium theory of awareness can account for similar hypotheses. Inasmuch the shape of the operative functions is known, it appears as possible to discuss the *directions* of the arising shiftings. It is remarkable that the detrimental and painful stimuli (work, energy loss, attacks against our organization etc.) have a predominant part in the formation of the new equilibrium. If detrimental stimuli appear, pain and unpleasantness will increase, while the awareness of the merely objective information (light, sound etc.) will decrease. It can be shown that the direction in the shifting of the optimum situation is a mathematical consequence. Objective sensation pertain to the first type, pains and unpleasant sensations to the second type of psychophysical functions.

Special cases arise when several detrimental stimuli appear simultaneously. It can be shown that they become competetive among themselves. We become the *most* aware of the *most* detrimental influences.

b.) *Adaptation.* Adaptation results in a decrease of differential sensitivity for the stimulus concerned. In the situation of complete adaptation (example: continuous presence of a clock with its ticking), differential sensitivity vanishes over a large and entire range. Therefore *sensation* itself is on a zero or on a stationary level within this range.

c.) *Change in the scope of the sensory awareness.* Owing to outside influences (fatigue, narcotics), also through our own concentration on mental activities, the total scope open for sensory awareness can decrease. The contrary happens in case a person uses "stimulatives", *i.e.* artificial expedients just to increase the total scope for receiving impressions and information from outside. In all similar situations the conditions for maximization will change accordingly and the equilibrium will be shifted.

3.) *The Time Integral of our Sensations.* In dynamical Psychology the intensities of our sensations are considered as variables in the flow of time. We arrive at the important notion of the Time Integral of the Sensations.

$$\int_0^t I \, dt$$

This Integral covers the idea of the accumulated sensations in the

flow of the time. The Integral seems to play a very important part in the operation of memory and oblivion. Namely, our memory and our education is built primarily on accumulated sensations.

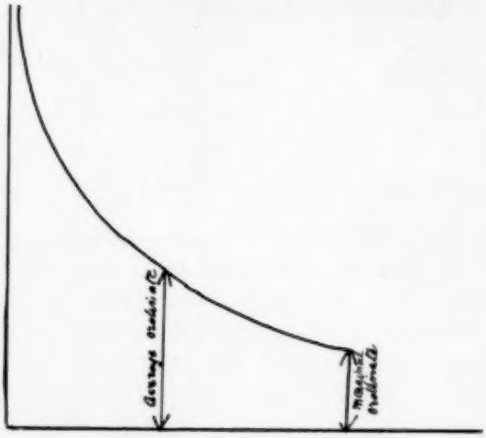


Figure II
The first psychophysical function

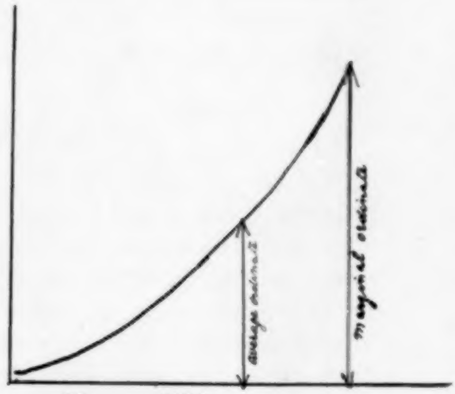


Figure III
The second psychophysical function

NOTE I.

Selection Problems of the Second Order.

Our previous hypothesis was built on the assumption that commodities were at our disposal in *abundant* quantities. The only restriction which we met was the *limitation of the volume*. Our hypothesis can be modified. We can assume the simultaneous existence of two types of operative restrictions, namely the *shortage of the available commodities* and the *limitation of the volume*. Several special cases can be constructed. The following are typical situations.

1.) The case of *relative abundance*. The available quantities of commodities do not actually cover our needs. According to the formulation of the marginal utility theory: the level of the marginal utilities is *not* zero. However, we have no volume sufficient to locate all units.

2.) *The case of sufficiency*. This is a borderline situation in which we dispose of the volume which is just sufficient to locate all commodity units.

3.) *The case of overwhelming commodity shortage*. The admitted volume for a certain type of commodity can be larger than needed to allocate all of its units. No selection is operative in view of this type of commodity.

4.) *Mixed cases*. Certain types of commodities are in abundance, for certain types we meet relative abundance, sufficiency or overwhelming shortage.

All the above problems might be classified as *optimum selection problems of the second order.* The conditions of maximization will accordingly be modified in view of the superimposed restrictions or in view of the alleviation of the limitations. Minimum problems will also appear in cases where relative, *i.e.* non-absolute abundance of the commodities is involved. The total utility of the commodities not included in the selection should be a *minimum.*

NOTE II.

Note about measurability.

In the deeper theory of Optimum Selection the direct, numerical measurability of utility is given up. It is sufficient to adopt *Vilfredo Pareto's* conception about the existence of an *indifference map.* Each curve on the indifference map is the locus of baskets of commodities to which a person attributes the same level of utility. Owing to the assumptions the indifference curves do not intersect. (See Figure I). Each curve can be marked with a "choice indicator" expressing merely the increase or the decrease of the utility levels. It can be shown that for the purpose of *maximization* the knowledge of the *topological order* of the choice indicators is sufficient, their numerical values are irrelevant.

If the conception of the indifference map is adopted, the operation of optimum selection can be visualized on the map by drawing the "*selection line.*" The selection line is the locus of all selected sets of commodities compatible with the volume-restriction. The *linearity* of the selection line covers the simplest hypothesis, *i.e.* the case in which the volumes occupied by the commodities are directly proportional to the quantities of commodities concerned. Optimum selection is materialized at the point at which the selection line is a *tangent* to one of the indifference curves.

NOTE III.

Exact formulation of the psychophysical functions.

The psychophysical functions can be formulated in an exact form on the basis of our experimental data concerning the *just noticeable differences,* the J.N.D.-s. As an illustration we shall discuss the derivation of the third psychophysical function:

The absolute threshold of a certain stimulus can be tested experimentally. It is the minimum stimulus (ν) which precedes *one* unit of a sensation in terms of J.N.D.-s. Also we have experimental information about the value of the so-called *Weber fraction* over a certain range of the stimulus. This fraction is supposed to be *constant* over a large middle range of the stimuli:

$$\text{WEBER FRACTION} = \frac{\text{stimulus increment producing one J.N.D.}}{\text{operative stimulus}} = \gamma = \text{const.}$$

The following table is supposed to give a general survey of the pro-

gression of stimuli and of sensations. Stimuli must be increased in terms of a *geometrical* progression in order to create the increase of sensation in the *arithmetic* progression. Sensations are measured in terms of J.N.D.-s. J.N.D.-es are simply *counted.*

Stimulus	*Number of J.N.D.-s.*
ν	0
$\nu(1+\gamma)$	1
$\nu(1+\gamma)^2$	2
etc.	

A continuous function between stimulus and sensation can be formulated and all the derived functions established.

NOTE IV.

It is obvious that

$$\text{Const.} = -a \ln \nu$$

This is the value of the integration constant which satisfies the requirement, "threshold stimulus, zero sensation."

$$I = a \ln \nu - a \ln \nu = 0$$

NOTE V.

Through reasoning by the means of the analogy the following considerations would follow. "The marginal utilities pertaining to the marginal volume elements are equal." (See p. 177) Therefore, in the equilibrium situation of awareness *the marginal rates of sensitivity pertaining to the marginal J.N.D.-s over all the operative stimuli are equal.*

It is outside the scope of this study to analyze the possibility and to interpret the meaning of a similar theorem.

THE DEPARTMENT OF MATHEMATICS

UNIVERSITY OF CALIFORNIA - LOS ANGELES

announces an

INSTITUTE FOR TEACHERS OF MATHEMATICS

July 9 - 20, 1951
on the

LOS ANGELES CAMPUS - UNIVERSITY OF CALIFORNIA

DIRECTOR: W. W. Rankin, Professor of Mathematics, Duke University

Two lectures daily will be given by outstanding men from business, industry, and the teaching profession. Study groups will be conducted daily by leaders of recognized ability chosen from teachers in the various fields of mathematical instruction.

Certificates of attendance will be issued to those registered in the Institute. It will also be possible for registrants to earn two units of college credit by attending lectures and by participating in a sufficient number of the study groups.

Application blanks for the Institute and further details may be obtained by writing to

Clifford Bell
Mathematics Department
University of California
Los Angeles, 24, California

THE ASSOCIATION OF TEACHERS OF MATHEMATICS IN NEW ENGLAND

will sponsor

THE INSTITUTE FOR TEACHERS OF MATHEMATICS

August 23 - 30

at CONNECTICUT COLLEGE, NEW LONDON, CONN.

If you wish a program, please drop a card to

Charles H. Mergendahl
Newton High School
Newtonville, 60, Mass.

or

Katherine F. O'Brien
Deering High School
Portland, Maine

A GENERAL ISOMORPHISM THEOREM FOR FACTOR GROUPS

Irving S. Reed

This note is concerned with the generalization of the Zassenhaus isomorphism theorem [1], [2], and its application to a theorem of Remak [3]. For this purpose we utilize the following basic terminology and propositions: $\stackrel{\sim}{\rightarrow}$ denotes a homomorphic mapping of a group onto a group and $\simeq$, an isomorphic mapping. (A) If $G \stackrel{\sim}{\rightarrow} \overline{G}$, there exists a kernel E (the set of elements in G which map onto $\overline{e}$, the unit of $\overline{G}$), a normal subgroup of $\overline{G}$, such that $G \stackrel{\sim}{\rightarrow} G/E$ [4]. If the mapping in (A) is designated by σ_E then $\sigma_E G = \overline{G}$ and $\sigma_E^{-1}\overline{G} = G$ ($\sigma_E^{-1}\overline{G}$, denoting the complete inverse image of $\overline{G}$). (B) If U is a subgroup of G (not necessarily proper) and $G \stackrel{\sim}{\rightarrow} \overline{G}$ with kernel E, then $U \stackrel{\sim}{\rightarrow} \overline{U} = \sigma_E U$ and $\sigma_E^{-1}\overline{U} = EU$; moreover, the kernel of $U \stackrel{\sim}{\rightarrow} \overline{U}$ is $E \cap U$. (C) If $U \subseteq G$ and $G \stackrel{\sim}{\rightarrow} \overline{G}$ with kernel E, then $U \stackrel{\sim}{\rightarrow} \overline{U} = \sigma_E U$; moreover, if $\overline{N}$ is a normal subgroup of $\overline{U}$, then $\sigma_E^{-1}\overline{U}/\sigma_E^{-1}\overline{N} \simeq \overline{U}/\overline{N}$.

THEOREM 1. Given $X_i \subseteq U_i \subseteq G$ $(i = 1, 2, 3 \dots n)$, U_i subgroups of the group G and X_i normal in U_i, then the following $n + 1$ factor groups are isomorphic:

$$\frac{U_1 \cap U_2 \cap U_3 \cap \dots \cap U_n}{(X_1 \cap U_2 \cap \dots U_n)(U_1 \cap X_2 \cap U_3 \cap \dots U_n)\dots(U_1 \cap \dots U_{n-1} \cap X_n)} \simeq$$

$$\frac{X_1(U_1 \cap U_2 \cap \dots U_n)}{X_1(U_1 \cap X_2 \cap U_3 \cap \dots \cap U_n)\dots(U_1 \cap \dots U_{n-1} \cap X_n)} \simeq$$

$$\frac{X_2(U_1 \cap U_2 \cap \dots U_n)}{X_2(X_1 \cap U_1 \ \dots U_n)(U_1 \cap U_2 \cap X_3 \cap U_4 \cap \dots U_n)\dots(U_1 \cap \dots U_{n-1} \cap X_n)} \simeq \dots \simeq$$

$$\frac{X_n(U_1 \cap U_2 \cap \dots U_n)}{X_n(X_1 \cap U_1 \cap \dots U_n)\dots(U_1 \cap \dots \cap U_{n-1} \cap X_{n-1} \cap U_n)}.$$

PROOF. Let $U_1 \stackrel{\sim}{\rightarrow} \overline{U}_1$ with kernel X_1 (the natural mapping). Then $\sigma_{X_1} U_1 = \overline{U}_1$. Let $\sigma_{X_1}(U_1 \cap U_2 \cap \dots U_n) = K_1$, and $\sigma_{X_1}[(U_1 \cap X_2 \cap U_3 \cap \dots \cap U_n)\dots(U_1 \cap \dots U_{n-1} \cap X_n)] = H_1$. Hence under this mapping

$$(U_1 \cap U_2 \cap \dots U_n) \stackrel{\sim}{\rightarrow} K_1, \tag{1}$$

$$(U_1 \cap X_2 \cap U_3 \cap \dots \cap U_n)\dots(U_1 \cap \dots U_{n-1} \cap X_n) \stackrel{\sim}{\rightarrow} H_1.$$

Since $(U_1 \cap X_2 \cap U_3 \cap \dots \cap U_n)\dots(U_1 \cap \dots U_{n-1} \cap X_n)$ is normal in

$(U_1 \cap U_2 \cap \ldots U_n)$, it is clear that H_1 is normal in K_1. By (B) we have

(2) $\sigma_{X_1}^{-1} K_1 = X_1(U_1 \cap U_2 \cap \ldots U_n)$ and

$$\sigma_X^{-1} H_1 = X_1(U_1 \cap X_2 \cap U_3 \cap \ldots U_n)\ldots(U_1 \cap \ldots U_{n-1} \cap X_n),$$

and that the kernel of the first mapping of (1) is

$$X_1 \cap U_1 \cap U_2 \cap \ldots U_n = X_1 \cap U_2 \cap \ldots U_n.$$

The complete inverse image of H_1 under this kernel is given by

(3) $\sigma^{-1}_{X_1 \cap U_2 \cap \ldots U_n} H_1 = (X_1 \cap U_2 \cap \ldots U_n)(U_1 \cap X_2 \cap \ldots U_n)\ldots$

$$\ldots(U_1 \cap \ldots U_{n-1} \cap X_n).$$

Now by (C) we have

$$K_1/H_1 \simeq \sigma_{X_1}^{-1}K_1/\sigma_{X_1}^{-1}H_1 \simeq \sigma^{-1}_{X_1 \cap U_2 \cap \cdots U_n} K_1/\sigma^{-1}_{X_1 \cap U_2 \cap \cdots U_n} H_1,$$

and by (2) and (3) that the first two factor groups of the theorem are isomorphic. The remainder of the theorem follows by replacing 1 by 2, 2, 4, ...n, successively. Hence theorem is proved.

The following corollary of Theorem 1 will be useful for the next theorem.

COROLLARY. If N and M are normal subroups of G and U is a subgroup of G, then

$$U/(U \cap N)(U \cap M) \simeq NU/N(U \cap M) \simeq MU/M(U \cap N).$$

PROOF. For $n = 3$ in Theorem 1 set $U_1 = U$, $U_2 = U_3 = G$, $X_1 = e$, $X_2 = N$ and $X_3 = M$.

THEOREM 2. If N_1, N_2, and N_3 are normal subgroups of G, then

$$\frac{N_1N_2 \cap N_1N_3 \cap N_2N_3}{(N_1N_2 \cap N_3)(N_1 \cap N_2)} \simeq \frac{N_1N_2 \cap N_1N_3 \cap N_2N_3}{(N_1N_3 \cap N_2)(N_1 \cap N_3)} \simeq \frac{N_1N_2 \cap N_1N_3 \cap N_2N_3}{(N_2N_3 \cap N_1)(N_2 \cap N_3)} \simeq$$

$$\simeq \frac{N_1N_2 \cap N_3}{(N_1 \cap N_3)(N_2 \cap N_3)} \simeq \frac{N_1N_3 \cap N_2}{(N_1 \cap N_2)(N_2 \cap N_3)} \simeq \frac{N_2N_3 \cap N_1}{(N_1 \cap N_2)(N_1 \cap N_3)} \simeq$$

$$\simeq \frac{N_1(N_1N_2 \cap N_3)}{N_1(N_2 \cap N_3)} \simeq \frac{N_2(N_1N_2 \cap N_3)}{N_2(N_1 \cap N_3)} \simeq \frac{N_1(N_1N_3 \cap N_2)}{N_1(N_2 \cap N_3)} \simeq$$

$$\simeq \frac{N_3(N_1N_3 \cap N_2)}{N_3(N_1 \cap N_2)} \simeq \frac{N_2(N_2N_3 \cap N_1)}{N_2(N_1 \cap N_2)} \simeq \frac{N_3(N_2N_3 \cap N_1)}{N_3(N_2 \cap N_1)} \simeq$$

are the twelve factor groups, connecting N_1, N_2, and N_3.

PROOF. In the corollary of Theorem 1 let U be respectively $N_1N_3 \cap N_2$,

$N_2N_3 \cap N_1$, and $N_1N_2 \cap N_3$, then

$$(4) \quad \frac{(N_1N_2\cap N_3)(N_1N_3\cap N_2)}{(N_1N_2\cap N_3)(N_1\cap N_2)} \simeq \frac{(N_2N_3\cap N_1)(N_1N_3\cap N_2)}{(N_2N_3\cap N_1)(N_2\cap N_3)} \simeq \frac{N_1N_3\cap N_2}{(N_1\cap N_2)(N_2\cap N_3)},$$

$$\frac{(N_1N_2\cap N_3)(N_2N_3\cap N_1)}{(N_1N_2\cap N_3)(N_1\cap N_2)} \simeq \frac{(N_1N_3\cap N_2)(N_2N_3\cap N_1)}{(N_1N_3\cap N_2)(N_1\cap N_3)} \simeq \frac{N_2N_3\cap N_1}{(N_1\cap N_2)(N_1\cap N_3)},$$

$$\frac{(N_1N_3\cap N_2)(N_1N_2\cap N_3)}{(N_1N_3\cap N_2)(N_1\cap N_3)} \simeq \frac{(N_2N_3\cap N_1)(N_1N_2\cap N_3)}{(N_2N_3\cap N_1)(N_2\cap N_3)} \simeq \frac{N_1N_2\cap N_3}{(N_1\cap N_3)(N_2\cap N_3)}.$$

From the identity,

$$(N_1N_2\cap N_3)(N_1N_3\cap N_2) = (N_1N_2\cap N_3)(N_2N_3\cap N_1) =$$

$$(N_1N_3\cap N_2)(N_2N_3\cap N_1) = N_1N_2\cap N_1N_3\cap N_2N_3,$$

we have by (4) that the first six factor groups of the theorem are isomorphic. For the last six factor groups let $U_1 = N_1N_2$, $U_2 = N_3$, $U_3 = G$, $X_1 = N_2$, $X_2 = e$ and $X_3 = N_1$ in Theorem 1 for $n = 3$, obtaining,

$$\frac{N_1N_2\cap N_3}{(N_2\cap N_3)(N_1\cap N_3)} \simeq \frac{N_1(N_1N_2\cap N_3)}{N_1(N_2\cap N_3)} \simeq \frac{N_2(N_1N_2\cap N_3)}{N_2(N_1\cap N_3)}.$$

Now permute N_1, N_2 and N_3 and the theorem follows.

It may be mentioned that Theorem 1 for $n = 2$ is due to Zassenhaus and that it was used to obtain an explicit construction for the equivalent refinements [5, page 6] of two normal chains. Theorem 1 for $n > 2$ leads naturally to a more general notion of refinement for a normal chain. A normal chain will be said to possess a lattice refinement if it is a lattice under group inclusion and every chain of the lattice is a refinement of the original normal chain.

To illustrate the way in which a lattice refinement arises, let $M_1 = M_1^{(t)} \supseteq \ldots \supseteq M_{s_t+1}^{(t)} = e$ $(t = 1, 2, 3)$ be three normal chains (n chains may be considered in a similar way). Define $M_{i,j,k}^{(1)} = M_{i+1}^{(1)}(M_i^{(1)} \cap M_j^{(2)} \cap M_k^{(3)})$ for $i = 1,\ldots, s_1 + 1$, $j = 1,\ldots, s_2 + 1$, $k = 1,\ldots,s_3 + 1$, $M_{s_1+1,1,1} = e$. Then $M_{i,s_2+1,k}^{(1)} = M_{i,j,s_3+1}^{(1)} = M_{i+1,1,1}^{(1)}$. Similarly, construct $M_{i,j,k}^{(2)} = M_{j+1}^{(2)}(M_i^{(1)} \cap M_j^{(2)} \cap M_k^{(3)})$ for $M_{1,1,s_2+1}^{(2)} = e$, and $M_{i,j,k}^{(3)} = M_{k+1}^{(3)}(M_i^{(1)} \cap M_j^{(2)} \cap M_k^{(3)})$ for $M_{1,1,s_3+1}^{(3)} = e$. In this way we we have constructed three lattice refinements of three normal chains, simultaneously, each of $s_1s_2s_3 + 1$ formally distinct terms. From

Theorem 1 for $n = 3$ they may be said to be equivalent in the sense, that

$$M^{(1)}_{i,j,k}/M^{(1)}_{i,j+1,k}M^{(1)}_{i,j,k+1} \simeq M^{(2)}_{i,j,k}/M^{(2)}_{i+1,j,k}M^{(2)}_{i,j,k+1} \simeq$$

$$M^{(3)}_{i,j,k}/M^{(3)}_{i+1,j,k}M^{(3)}_{i,j+1,k}.$$

In conclusion it may be noted that the above results apply in the usual way to ideal theory.

REFERENCES

1. H. Zassenhaus, Zum Satz von Jordan-Holder-Shreier, Abh. Math. Sem. Univ. Hamburg, v. 10 (1934) pp. 106-108.
2. H. Zassenhaus, Lehrbuch der Gruppen-Theorie, Leipzig, (1937).
3. S. Remak, Uber Untergruppen direkter Produkten von drei Faktoren, Jour. f. Math. (1931) pp. 65-100.
4. L. Pontrjagin, Topological Groups, Princeton (1946).
5. N. Jacobson, The Theory of Rings, Amer. Math. Soc. (1943).

California Institute of Technology

COLLEGIATE ARTICLES

Graduate training not required for reading

COMPLEX GRAPHS

Ed Bergdal

In a regular Cartesian bi-axial graph any point represents a pair of co-ordinate numbers, both of which are real numbers. Hence, as long as we confine ourselves to all-real co-ordinate values of a function, we can plot any points that we may deem to be necessary or convenient to use, and we can construct a curve through such points, which curve will graphically represent the function throughout the scope of its all-real values. Such a curve is shown in the upper part of figure 1. However, a regular Cartesian graph does not provide for the representation of complex or imaginary values.

It would seem at a first glance that if we want to represent a complex number in a bi-axial graph we must either dispense with any representation of a paired real number, and let a point represent a complex number only, or we must find a way to represent a pair of co-ordinate numbers otherwise than by a point. The first alternative is used in the Wessel-Argand-Gauss geometrical representation of complex numbers, which, because it requires two axes to represent a single quantity, can not serve as a graph of a function. Apparently if we are to succeed in constructing a functional graph, and at the same time retain the orthodox conception of imaginary numbers, we must use the second alternative.

Edmond Nicolas Laguerre developed an idea of a graphical representation of a complex number as a line segment. This is sometimes called "Laguerre's realization of complex numbers." Laguerre, however, does not provide any actual graphs. Laguerre depends on various abstract conceptions, such as circles with diameters approaching zero and points located at infinity, which conceptions are difficult to associate with such a relatively concrete thing as a graph. The complexities involved in Laguerre's procedure is doubtless what has prevented this procedure from being applied to practical use and has turned it into an almost forgotten mathematical curiosity*.

The writer developed independently - being at the time unacquainted with Laguerre's work - a simpler approach, by which it is assumed that when in a function y represents a real value and x a complex value, as $a + bi$, for every value of y there are two simultaneous values of

*Laguerre's discussion of the subject, *Sur l'emploi des imaginaires en Géométrie*, was published in *Nouvelles Annales de Mathématiques* in 1870, and was later (1905) republished in volume 2 of Laguerre's collected works, *Oeuvres de Laguerre*.

x, namely $a + b$ and $a - b$; thus for each ordinate there would be a pair of abscissas, giving a *pair of points* in the graph. These points correspond in effect with the limiting points of Laguerre's line segments. Such a pair of points must be considered to be a single graphical element, and the individual points constituting a pair can not be considered separately. Curves drawn through such points result in the same kind of graphs as those presented in this paper, tho these are produced differently. The basic principle on which this procedure is based is the assumption that an imaginary number is both positive and negative simultaneously. The meaning of this is admittedly obscure, but fundamentally not any more so than that of the equation $i^2 = -1$.

However, it would admittedly be desirable to be able to avoid the awkward neceesity for using either a line segment or a pair of points to represent a single pair of co-ordinate values in a graph; in fact, the incongruity of this indicates that tho we may be traveling in the right general direction, we are not on the main trail. We shall see that if we are willing to modify slightly our conception of the nature of imaginary numbers, with no sacrifice of consistency or of results obtained in the orthodox manner, we shall be able to represent a pair of co-ordinate values in a graph by a single point even when complex numbers are involved.

Algebraic operations indicate that square roots and other even roots of negative numbers exist, and they can not be ignored. When we say that an imaginary number multiplied by itself or by another imaginary number gives a negative product we have hitherto assumed that the operation is "real", or regular, but that the numbers are "imaginary", that is, they are irregular in that they are of intangible values which can not be expressed in terms of real numbers. We find that proceeding under this assumption leads to self-consistent and useful results.

We should bear in mind, however, that such intangible imaginary numbers are merely assumed to exist. It happens that we are not dependent on this assumption for the results sought. Instead of assuming that we have intangible imaginary numbers to which we apply real operations, we may assume that we apply imaginary operations to real numbers, such operations being imaginary in the mathematical sense, so that the result of any application of an imaginary operation to a real number would be identical with the result of a corresponding application of a real operation to an imaginary number. Thus, the real number 1 squared imaginarily would equal -1. We can therefore let the symbol i represent a real number 1 which when squared must be squared imaginarily. The nature of the operation of squaring imaginarily may be obscure, but not any more so than the nature of the orthodox imaginary numbers. However, we shall examine this question further a little later. Since i^3 equals $-i$, it must therefore equal -1 if the results are to conform to those obtained with the orthodox imaginary-number system. It will be

seen that the values of the successive powers of i would be as follows: $i = 1$, $i^2 = -1$, $i^3 = -1$, $i^4 = 1$, $i^5 = 1$, $i^6 = -1$, $i^7 = -1$, and so on, pairs of positives alternating with pairs of negatives. It may be noted that the term "power" as used in this paper means the indicated quantity in question, and not the exponent.

The acceptance of the assumption that the operations in question are imaginary does not necessarily preclude the designation of i as an imaginary number, for we may define an imaginary number as a number to which imaginary operations are to be applied. In spite of the fact that we assume such a number to be real, the designation "real number" will hereinafter, to avoid confusion, be reserved for such numbers as are ordinarily called real, for regardless of reality, it is necessary to distinguish numbers to which special operations are to be applied from other numbers.

We may "realize" the even powers of i when they appear, that is, we may dispense with the i-notation and we may add these numbers to, or subtract them from, other numbers, but we may not do this with the odd powers of i until we are thru with all operations which involve shifts in the powers of i. But at the end of our operations we can do it, so that, for instance, $2 + i$ becomes 3, and $3 - 4i^3$ becomes 7. The values thus obtained we may for convenience call the *static values* of imaginary or complex numbers. The static value of an imaginary number is the same as the static value of a corresponding real number, but what we may call *dynamic* or *operational values* of the two numbers are different. However, these terms are used for convenience only, being tentatively adopted in the absence of anything better; what are here called static value and dynamic value are perhaps more properly speaking merely separate manifestations of a single value. The assigning of static values to imaginary numbers does not involve any inconsistency with the orthodox imaginary-number system; it merely amplifies the operational possibilities.

It will be seen that the method of operating is exactly the same whether we proceed under the orthodox assumptions or under the new assumptions whenever the operations cover the same ground. Hence, we may, if we want to do so, retain the orthodox assumptions except that at the end of the operations we may assign the above-indicated static values to any imaginaries that may be left uncanceled. But we must remember that when we have found a root of an equation we are not at the end of our operations if we are to substitute the root for an unknown in the equation, and since a root can be put back into the equation only in the form it was taken out of it, we may not realize any imaginaries in such root. This, however, does not prevent us from using the static values for the purpose of plotting points representing such roots in a graph, for in such a case we are at the end of our power-shifting operations as far as the construction of the graph is concerned.

It follows that if we assume that -1 has a positive square root, we may also consistently assume that it has a negative square root, but if

we define i as the positive root, $-i$ becomes the corresponding negative root, and there will be no confusion.

It is apparent that if we assume that we are dealing with real numbers and imaginary operations, we are in a position to inquire into the nature of such operations. We know that the raising of a number to a higher power is an operation of multiplication. If we want to multiply 1 so as to get -1 as a product we must evidently use -1 as a multiplier. Also, since i^2 and i^3 both equal -1, when we want to raise i from the second power to the third power we must use a multiplier which will not change the sign, a positive multiplier, that is, the number 1. In the operation leading to the next-higher power, i^4, which power equals 1, the sign of the product must be changed again, however; therefore our next multiplier must be -1. Hence, successive powers and static values of i are produced as follows:

$$i = 1$$
$$i^2 = 1(-1) = -1$$
$$i^3 = 1(-1)1 = -1$$
$$i^4 = 1(-1)1(-1) = 1$$
$$i^5 = 1(-1)1(-1)1 = 1$$
$$i^6 = 1(-1)1(-1)1(-1) = -1$$
$$i^7 = 1(-1)1(-1)1(-1)1 = -1$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

It is immaterial as far as the effect is concerned whether we assume, for instance, that i^3 is actually the product of 1(-1)1, or assume that it is produced by an intangible operation giving the same result; the important thing is the result. In ordinary powers we have a sequence of repeated factors which may all be positive, as, $a\cdot a\cdot a\cdot a\cdot a\cdot a = a^6$, or all negative, as, $(-a)(-a)(-a)(-a)(-a)(-a) = a^6$, while in the powers of imaginary numbers we have, or have the effect of, a sequence of factors which alternate between positive and negative, as indicated above. When we use a repeated positive factor the powers are all positive numbers, and when we use a repeated negative factor the powers alternate between one positive and one negative, while when we use factors which themselves alternate between positive and negative, the signs of the powers alternate in pairs of two positives and two negatives.

We have habitually assumed that the shifting of powers of numbers is something which can be performed in only one way; hence, when indicated results have not conformed to regularly-expected results we have assumed that the irregularity must lie in the numbers, and not in the operations. In the foregoing we have called attention to the fact that we may in such cases assume that the operations are imaginary in the mathematical sense, which merely means that we assume them to be

irregular, and which assumption leaves us with imaginaries having tangible values. A striking example of the use of the latter assumption is its application to graphs, a demonstration of which is the principal object of this paper.

In the accompanying figures, the upper part of figure 1 shows a regular Cartesian graph of the function $y = x^2 - 8x + 19$. Since this curve does not cross the X axis, the roots of the equation $f(x) = 0$ are evidently complex. Thru other operations we may find that these roots are $4 + \sqrt{-3}$ and $4 - \sqrt{-3}$. We see that the real part, 4, of the roots is the same as the abscissa of the vertex and axis of the curve, which happens to be a parabola. This is true for any complex value of this function. This relationship holds for any parabolic curve having a vertical axis. Hence, in computing a table of co-ordinates in which x has complex values, we may take for such complex values the constant real part and any convenient values for the imaginary part. The resulting complex numbers represent values of x only; to get the corresponding values of y we must substitute the selected complex values for x in the equation. We may note that for every value of y in this function there are two values of x, regardless of whether x is real or complex. The following table gives some of the co-ordinate values, with complex values of x, for the function $y = x^2 - 8x + 19$:

x	$4-4i$ $=0$	$4-3i$ $=1$	$4-2i$ $=2$	$4-i$ $=3$	4 $=4$	$4+i$ $=5$	$4+2i$ $=6$	$4+3i$ $=7$	$4+4i$ $=8$
y	-13	-6	-1	2	3	2	-1	-6	-13

The points represented in this table are shown plotted in the lower part of figure 1. If we draw a curve thru these points we find that we have another parabola identical in form with the first, the two curves having a common vertex, but while the "real" curve extends upward from the vertex, the complex curve extends downward. The knowledge of this enables us, after we have constructed a real parabola, to construct a corresponding complex parabola without computing any co-ordinate values. A curve based on the above points is shown in the lower part of figure 2.

When a complex value of x is to be derived from such a graph it is necessary to scale off the imaginary part from the abscissa of the real part. Hence, a line representing this abscissa, which, as we have seen, is constant, is also needed. This median line distinguishes a complex curve from a real curve occupying the same position. It is shown in figure 2 and in all the following figures.

Figure 3 represents an upper real curve and a lower complex curve of the function $y = x^2 - 8x + 20$. It is assumed that the lower curve is constructed by using the upper curve as a pattern. The lower curve crosses the X axis at the abscissas 2 and 6. Scaling these points off from the median line shows that they are equivalent to $4 - 2i$ and

$4 + 2i$ respectively, which are the roots of the equation $f(x) = 0$.

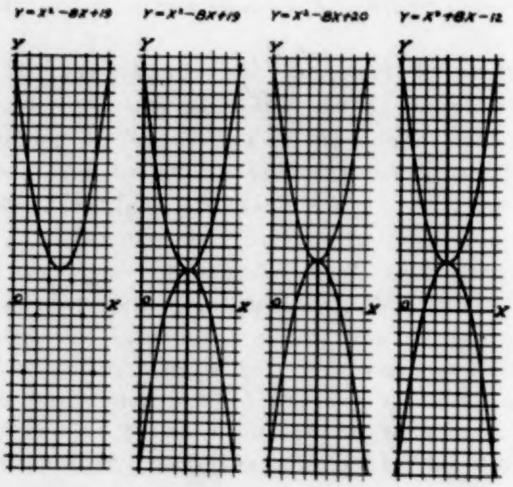


Fig. 1 *Fig. 2* *Fig. 3* *Fig. 4*

It may be noted that the curves we have considered come in pairs of one real and one complex, united at a common vertex. Such a pair may be considered to be a single complete graph having real and complex component parts. This applies not only when the zeros of $f(x)$ are complex but also when a function is one which is not ardinarily thought of as involving imaginaries, such as that represented in figure 4. When the real part crosses the X axis obviously the complex part does not, and vice versa. This is illustrated in figures 3 and 4 which show two complete graphs occupying the same position, but one having the complex portion at the bottom and the other having it at the top.

We thus see that we can construct a complex graph which does everything that any other functional graph does, in fact more. The curve has a characteristic form for each function; each co-ordinate set of values for x and y has a definite graphical representation, and each such representation corresponds to a definite set of co-ordinate values and to no other; and the abscissas of the points of intersection with the X axis represent the roots of the equation $f(x) = 0$. Also, beyond the vertex of the complex curve the graph may be continued as a real curve, or this process may be reversed, thus enabling one to show both complex and real values of x or y on the same graph. In addition, there is provided a graphical demonstration of the fact that complex values are involved even in functions in which the roots of the equation $f(x) = 0$ are real.

We may also note that simultaneous equations with complex roots may be solved graphically. Since it is evident that the complex portions of the graphs of such equations must intercept each other, and that their median lines must lie along the same abscissa or ordinate, we know that when these conditions do not prevail, the equations have no common complex roots.

As we have seen, in the case of a parabola the complex curve is another parabola, reversed in position, but of the same form as the real parabola. In other cases, however, the forms of the curves are different. In developing complex curves in connection with other conics we find the facts illustrated in figures 5 to 9.

$16x^2-9y^2+32x+54y+79=0$

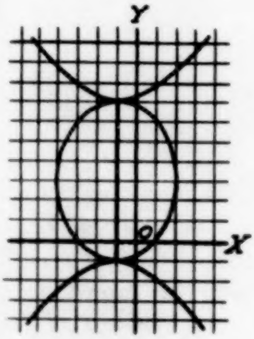


Fig. 5

$16x^2-9y^2+32x+54y-209=0$

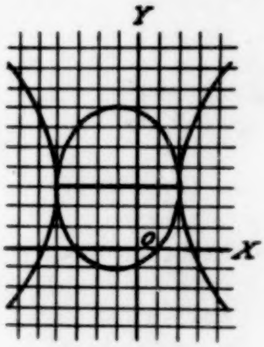


Fig. 6

$16x^2+9y^2+32x-54y-47=0$

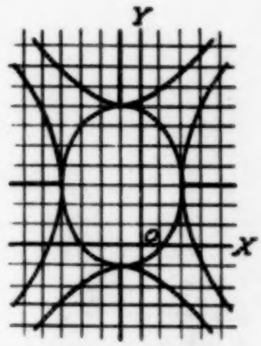


Fig. 7

Figure 5 shows a real hyperbola with the corresponding complex curve lying between the two branches of the real curve. The complex hyperbola does not have the form of a real hyperbola; it has the form of a real ellipse. It has complex values of x, as indicated by the vertical position of the median line.

Figure 6 shows a real hyperbola which is conjugate to that shown in figure 5. The complex portion, which occupies the same locus as the corresponding portion in figure 5, has complex values of y, and therefore has a horizontal median line. In a graph of this kind, where the median line is horizontal, the complex portion indicates complex roots of $f(y) = 0$, not of $f(x) = 0$. The complex roots of $f(y) = 0$ are found where the complex curve crosses the Y axis.

Figure 7 shows a real ellipse with complex curves on four sides of it. The latter have complex values of x at the top and bottom and such values of y at the sides. These values occur alternatively, not simultaneously. The complex curves have the form of a pair of conjugate hyperbolas. We may note that the real ellipse occupies the same locus as the complex curves shown in figures 5 and 6. This brings out the remarkable fact that a single locus may be occupied by three curves

representing three different equations and having three different sets of co-ordinates. The co-ordinates are different in that in one case they are all-real, in another they are complex-real, and in the third they are real-complex. The curves are distinguished from each other visually by the absence of the median line in one case, and by the different positions of such lines in the other cases.

$16x^2+9y^2+32x-54y+241=0$

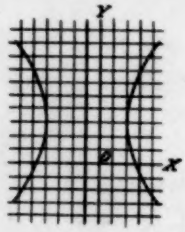


Fig. 8

$16x^2+9y^2+32x-54y+241=0$

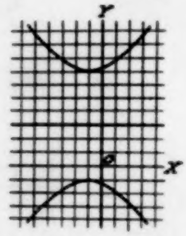


Fig. 9

In figures 8 and 9 we conjure forth one of those spook-functions which have been rapping for attention, but which have heretofore kept out of the light of day entirely, namely such functions as have no real locus, because they exist only as complex functions. The curve here presented is what we might call an all-complex ellipse. Figures 8 and 9 may be regarded as a single graph which could be presented in a single figure, but which would when so presented tend to be somewhat confusing to anyone who is not familiar with it. This is so because in these figures the median lines lie at right angles to what would normally be considered the respective axes of the portions of the curves involved, contrary to the situation existing in the preceding figures, where the median line coincides with the axis in each case. As indicated by the median lines, figures 8 and 9 represent complex values in x and y respectively.

Since a circle is in effect an equilateral ellipse, we may deduce that the corresponding complex curve should have the form of an equilateral hyperbola, which in fact is so.

In the foregoing we have considered complex values of x and y alternatively, but it is evident that we may readily plot complex values of x and y simultaneously if such a procedure should be called for.

Central Washington College of Education

GEOMETRY OF THE IMAGINARY TRIANGLE

O. J. Ramler

Definitions. A real straight line lying in the plane of a real circle and whose distance from the center of the circle is greater than the radius of the circle will be said to intersect the circle in a pair of imaginary points. These points together with any real point on the circle we shall define as an imaginary triangle inscribed in the circle. There are many theorems and properties of an ordinary real triangle inscribed in a circle which can readily be interpreted for, and extended to an imaginary triangle as defined above.

Centroid and Orthocenter. Since in any triangle the midpoint of a side is the foot of the perpendicular from the circumcenter O upon that side, we shall define the foot of the perpendicular from the center of a circle upon any line as the midpoint of the segment, real or imaginary, determined on the line by the circle. Thus for an imaginary triangle inscribed to a real circle a vertex A and the midpoint L of the opposite side are real points. The line joining them is the only real median this triangle has, and the point G on it which divides it from the vertex in the ratio 2:1 is defined as the centroid of the triangle. The imaginary triangle has one real altitude AD, namely the perpendicular from A to the opposite side l. Now in any triangle inscribed in a circle the segment of an altitude from the orthocenter to the side equals its extension from the side to the circumcircle(1). We use this property to define the orthocenter H, of an imaginary triangle. In an imaginary triangle (A, l) the reflection H in the real line l of the point D' other than the vertex A where the real altitude AD cuts the circumcircle is defined to be the orthocenter H of the imaginary triangle.

The Nine-Point Circle. The circle whose center is the midpoint J, of the segment HO, and whose radius is one half the radius of the circumcircle (O), is defined to be the nine point circle of the imaginary triangle. The circle (J) is then directly homothetic to the circle (O) with respect to H in the ratio 1:2. Since D is the midpoint of HD', D lies on circle (J); moreover since AH and OL are parallel, $HJ = JO$, J is on the perpendicular bisector of DL; hence $JL = JD$, and L also lies on circle (J). Again, the midpoint P of HA also lies on circle J, thus we have identified as real points, three of the nine points usually associated with the nine point circle of any real triangle.

PDL is a right triangle inscribed to circle J, hence PJL is a straight line, and JL is parallel to and equal to one half OA. Thus the nine point circle (J) and the circumcircle (O) are homothetic in the ratio $-1/2$; but $GL:GA = -1/2$, hence G is a center of similitude. Hence G lies on OH, and OL equals one half of HA, as is the case for real triangles.

The Polar Circle. The polar circle of a triangle is the circle whose center is the orthocenter and whose radius is given by $r^2 = HA \cdot HD$ where H is the orthocenter, and AD is the altitude from vertex A. In other words r^2 is the power of H as to any circle having AD as a chord. For an imaginary triangle (A, l), H never lies between A and D, hence the polar circle of an imaginary triangle is always real, and the vertex A and the line l are pole and polar with respect to the polar circle. If P is the midpoint of HA, the nine point circle passes through P and D and its center is the midpoint of HO. Now the inverse of the circumcircle (O) with respect to the polar circle is a circle because for an imaginary triangle H cannot lie on or within the circumcircle. The center of the inverse circle lies on the line HO. Its radius R' is given by the formula

$$R' = \frac{Rr^2}{t},$$

where t is the power of H as to the circumcircle, R is the circumradius and r is the radius of the polar circle (circle of inversion)[2]. Now $r^2 = HA \cdot HD$ and t is $HD' \cdot HA$ where D' is the point in which AH meets the circumcircle. Hence $\dfrac{R \cdot HA \cdot HD}{HD' \cdot HA} = \dfrac{R}{2}$, since D is the midpoint of HD'. Moreover D is the inverse of A, and P is the inverse of D' as to the polar circle, hence the inverse of the circumcircle as to the polar circle is the nine point circle as is the case for real triangles. It follows that the circumcircle, the nine point circle and the polar circle are coaxal[3].

Isogonally Conjugate Points. Let us consider first a pair of points P and P' isogonally conjugate with respect to a real triangle ABC. If AP and AP' are produced to cut the circumcircle in E and E' respectively, then EE' and BC are parallel. It can readily be proved that if X is the point where PE' cuts BC, then XP' is parallel to AP. This follows at once from the fact that the points on AP and their isogonally conjugates on AP' are projectively related, and the line BC is the axis of homology. This property can be used to locate and define the point P' isogonal to P in an imaginary triangle (A, l). Choosing any point P, let AP cut the circumcircle (O) in E and line l in T. Let the line through E parallel to l cut circle (O) in E'. Let PE' cut l in X; through X draw a line parallel to AP cutting AE' in P'. We define P' to be isogonal to P as to the imaginary triangle (A, l). Since the isogonal of any point on the bisector of an angle is also a point on that bisector, the pairs of points mutually isogonal and lying on the bisector of an angle form an involution of points whose self corresponding points are the incenter and excenter lying on that bisector. When the triangle is imaginary, these double elements are also

imaginary since the involution is then of the elliptic type. Hence an imaginary triangle has no real in - and escribed circles.

Symmedian point and Brocard Circle. In a real triangle the centroid G and the symmedian point K are isogonally conjugate. From the construction given above we can locate a real point which is isogonally conjugate to the centroid G of an imaginary triangle. The point thus found is defined to be the symmedian point K of the imaginary triangle. The circle on OK as diameter may now be defined to be the Brocard circle of the imaginary triangle. Brocard's first triangle can now be defined and determined for the imaginary triangle (A, l). The Brocard circle is the circumcircle of the Brocard first triangle and since a triangle and its first Brocard triangle are similar, the first Brocard triangle of an imaginary triangle will also be imaginary. Its real vertex A' will be the intersection with the Brocard circle, other than O, of the perpendicular to l from O. Since a triangle and its Brocard triangle have the same centroid G, the real side l' of Brocard's first triangle can readily be found and defined. Extend $A'G$ to L' so that $A'G:GL'$ = 2:1. At L' erect a line l' perpendicular to $O'L'$ where O' is the midpoint of OK. Then the triangle determined by A', l' and circle (K) is Brocard's first triangle of the imaginary triangle (A, l). The Steiner and Tarry points can now be found. The Steiner point S is the end of the chord AS of the circumcircle, where AS is parallel to the side l' of Brocard's first triangle. The Tarry point T is diametrically opposite to the Steiner point on the circumcircle (O).

The Mean Triangle. With every triangle is associated its mean triangle as defined by P. Delens[4]. The mean triangle of ABC is the equilateral triangle inscribed to the circumcircle of ABC and whose vertices are the ends of the circumradii drawn parallel to and similarly directed to the cusp tangents of the Steiner Deltoid of the triangle ABC. One vertex of the mean triangle for the imaginary triangle (A, l) is then found as follows: Draw the line from the circumcenter O perpendicular to l cutting the circumcircle in U and V. Trisect the arc UA (or VA) at M so that $AM = 2MU$. Then M is one vertex of the mean triangle; the remaining two vertices are then uniquely determined since the mean triangle is always equilateral. Suppose now, that a side l' of a second triangle (real or imaginary) inscribed to circle (O) is given; the opposite vertex A' then can be found so that the two triangles (A, l) and (A', l') have the same mean. Through A draw a line parallel to l' cutting the circle (O) at X; draw the chord XA' parallel to l. Then (A', l') and (A, l) have the same mean triangle, and are mutually orthopolar[5].

Simson lines and Orthopoles. It is well known that if the perpendicular from a point P of the circumcircle to any side BC of a real triangle ABC is extended to meet the circumcircle at P', then AP'

is parallel to the Simson line of P for triangle ABC[6]. Thus the direction of the Simson line of a point is determined. The line through the foot of the perpendicular from P upon BC and having this direction is then the Simson line of P for the triangle ABC. The extension to an imaginary triangle (A, l) is obvious: drop a perpendicular PQ from P of the circumcircle (O) to line l, cutting the circumcircle again at P'; the line through Q parallel to AP' will be defined to be the Simson line of P for the imaginary triangle (A, l).

Since the orthopole[7] of a line cutting the circumcircle is the point of intersection of the Simson lines of the points where the line cuts the circle, we can readily define the orthopole of a line cutting the circumcircle of an imaginary triangle with respect to that triangle to be the intersection of the Simson lines as defined above. In case the line does not cut the circumcircle we can still define the orthopole of a line m with respect to an imaginary triangle by a construction based on the following theorem: If two triangles inscribed in the same circle have a common mean, the midpoint of the segment joining their orthocenters is the orthopole of any side of one triangle with respect to the other[5]. This suggests the following construction to find the orthopole of any line m with respect to an imaginary triangle (A, l): Let l' be the line whose orthopole with respect to (A, l) is wanted. Locate vertex A of the triangle (A', l') so that (A, l) and (A', l') have the same mean triangle. The midpoint of the line segment joining their orthocenters is the orthopole of l' with respect to triangle (A, l).

REFERENCES.

1. R. A. Johnson, Modern Geometry, §254.
2. R. A. Johnson, Modern Geometry, p. 50.
3. N. Altshiller-Court, College Geometry, p. 217.
4. P. Delens, Mathesis, Vol. 51, 1937, p. 264.
5. O. J. Ramler, American Mathematical Monthly, 1940, p. 140.
6. L. S. Shively, An Introduction to Modern Geometry, 5.9 p. 51.
7. R. A. Johnson, Modern Geometry, §406.

The Catholic University of America

TEACHING OF MATHEMATICS

Edited by

Joseph Seidlin and C. N. Shuster

This department is devoted to the teaching of mathematics. Thus articles on methodology, exposition, curriculum, tests and measurements, and any other topics related to teaching, are invited. Papers on any subject in which you, *as a teacher*, are interested, or questions which you would like others to discuss, should be sent to Joseph Seidlin, Alfred University, Alfred, New York.

TUTORIAL WORK IN MATHEMATICS AT THE UNIVERSITY OF BUFFALO*

Harriet F. Montague

Since 1931 the senior division of the College of Arts and Sciences at the University of Buffalo has been under the so-called Tutorial System whereby each student, having been accepted by a department as a majoring student, is given opportunity to pursue independent work in that particular subject. The object of this paper is to describe the tutorial system as it functions in the Department of Mathematics in the hope that some of its features may prove useful to departments of mathematics in other colleges even though they do not have the complete system.

Our philosophy with regard to majors in mathematics is consistent with the general philosophy of the tutorial system that each student should be given the opportunity for individual study and research. Although this can be done in some departments without a great deal of background we have felt that students in mathematics at the beginning of their junior year do not have an adequate background to begin individual projects. The average student at this time lacks a certain maturity in mathematics necessary for such projects. Our program of tutorial work in the junior year is devised to provide background material and develop a degree of mathematical maturity. The students are divided into small groups of three or four each. These groups meet once a week throughout the year with an instructor to study and discuss the history and foundations of mathematics and also the various branches of mathematics No text book is used and we try to avoid definite assignments. General topics are suggested by the tutor for discussion. The students then use the library to gather material to bring back the following week. Certain references are recommended but the students are urged to use additional books and periodicals for supplementary material. Topics chosen outside the history of mathematics include mathematical logic, the number system, types of geometries, fundamental concepts of the calculus. Much of the success of these discussion groups depends on the tutor. He must encourage the students to use the library intelligently. He must also encourage them to do some independent thinking. He must remain in the background during the discussion periods, but still guide the discussion so that it is fruitful and distributed throughout the group. It is an assignment which can be done in a mediocre way more easily than it can be done well if the tutor is not convinced

*This is a portion of a paper presented to the Upper New York State Section of the Mathematical Association of America, Syracuse, New York, April 22, 1950.

of its value.

The first half of the senior year is devoted to the investigation, solution, and writing-up of an individual problem. Here we reach the very heart of the tutorial system. It may seem an impossible task for all seniors - average, superior, and below average to accomplish this feat but it is done. Some times the tutor needs to come to the rescue but he tries to remain in the background as much as possible and to act merely in the role of an umpire. Again we have small groups, preferably not more than three in a group. Each student is supposed to find for himself a problem which he would like to work on. It might be a problem arising from a course he is taking or has taken. It might be a problem proposed for solution in one of the publications. The tutor helps him to decide if the problem is worthwhile. If it is too difficult it may be simplified, if too trivial it may be strengthened by extension or enlargement. When each person has decided on a problem he reports his progress to the group which meets once a week. The tutor and the other members of the group encourage or criticise him and he goes off for further progress. We feel that the writing-up of the problem is very important. It is excellent practice for the student going on to graduate work. For the student who will not go on it is an accomplishment he never otherwise would have achieved. Therefore, even if a problem has not been completely solved, the attempt is presented and written up.

The second half of the senior year is concerned with a synthesis of all the work taken in mathematics. At the end of this semester the students are required to take the comprehensive examination in mathematics. This last part of our tutorial work can hardly ignore that fact and there is to some extent a general review in preparation for the comprehensive. However, that is not its primary purpose. Its primary purpose is synthesis and correlation. Here the student can see how his various courses are related. He can compare techniques used in different courses. He can appreciate how one course helps another in the solution of a problem. He can observe the refinement and finesse which are used in more advanced courses. Some of our questions on the comprehensive are designed to test such things as these rather than to test just factual knowledge. The comprehensive examination has one part in which the student is confronted with material completely new to him. He is asked to read the material, perhaps supply some missing steps or verify some of the steps, and then answer some questions relating to the material presented.

With the proper attitude on the part of the tutors and the students there is a great deal of satisfaction and reward which comes as a result of this work in small groups. It is surprising what things come to be discussed at times in these sessions - things academic and things extra-curricular, personal problems and campus gossip. The tutor becomes the confidante and can exert a great influence on the students. Opportunity arises and should be welcomed for the tutors to discuss with the students employment opportunities and requirements for specific jobs. Other things such as the use of the library and the various mathematical journals can be discussed from time to time. We do not have opportunities to talk about these things in our regular courses as much as we should.

We believe our tutorial system in mathematics is a valuable one for the students and the tutors. We hope that it will continue to be a challenge to us and perhaps to others outside our own institution.

University of Buffalo

THE THEORY OF FUNCTIONS OF A REAL VARIABLE

John W. Green

Introduction. In this article, the theory of functions of a real variable will be considered from the narrowest possible point of view, as a theory set up to examine critically the processes of analysis and to extend and generalize these processes to as many as possible of the situations that arise in special branches of analysis, such as potential theory, differential equations, and trigonometric series. Actually it is difficult to fix precisely the boundaries of the subject, since to a considerable extent it overlaps with topology, logic and foundations, and the theory of abstract spaces. University courses in the subject are likely to vary extremely widely from one to the other in the subject matter presented and the point of view of the presentation. However, the accomplishment of the aim set forth above will usually occupy a central position, be it ever so disguised.

The theory of functions of a real variable is based on the theory of the real numbers; consequently, one of our first tasks is to look at the real numbers and determine just what properties they must be assumed to have in order to develop the theory. It will be assumed that the reader has an intuition for the rational numbers (integers and fractions; positive, negative, and zero) and is familiar with the rational operations of addition, multiplication, subtraction, and division, and the laws governing those. Also it is assumed that he is familiar with the notion of order (a less than b, written $a < b$) and the most common rules pertaining to it. In algebraic language, it is assumed that the reader knows that the rational numbers form an ordered Archimedian field. How this is known is a subject of its own and will not be gone into here. The existence of such a system may be postulated outright, or a system may be constructed by suitable definitions from the system of positive integers, whose existence and basic properties (very few in number) may be postulated outright, or various other approaches may be used. In this matter our field of study is closely connected with logic and foundations, and with algebra.

At any rate, let us assume that the rational numbers and their arithmetic are at our disposal, but that our intuition fails to provide us with any further numbers. Our system, then, is not sufficiently general to satisfy many of the demands of the processes of analysis. For example, the equation $x^2 = 2$ has no solution, and the expression $(1 + 1/n)^n$, $n = 1, 2, 3, \ldots$ does not have any limit as n increases, in spite of the fact that it always increases and remains less than 3. The question arises then: what further properties should the real numbers possess in order to eliminate the specific difficulties mentioned and a host of similar difficulties? It turns out that one property is needed; the real numbers should be complete, in the sense of

topology of metric spaces. This means the following. Let $x_1, x_2, \ldots$ be a succession (or sequence, as we shall call it) of real numbers such that to each positive real number ϵ there corresponds an integer N such that whenever n and m are greater than N, then $|x_n - x_m| < \epsilon$. Then there exists a real number x such that for every positive ϵ, all x_n lie within a distance ϵ of x from a certain subscript onwards. Now this is a very complicated condition, and its comprehension may require quite a little thought. However, in terms of language which the reader has probably encountered in the elementary calculus it reads as follows: Every sequence of real numbers which satisfies the Cauchy convergence criterion (or shorter, every Cauchy-convergent sequence) has a limit.

As examples, consider the two sequences: (a) 1, $1 + \frac{1}{2}$, $1 + \frac{1}{2} + \frac{1}{4}, \ldots$, and ($b$) 1, $1 + \frac{1}{2!}$, $1 + \frac{1}{2!} + \frac{1}{3!}$, $\ldots$. The sequence (a) is easily seen to have the limit 2; in fact $|2 - x_n| = 1/2^{n-1}$. In the case of (b) it is not so easy to see whether there is any limit. However, using the fact that $n! = 1 \cdot 2 \cdots n \geq 2^{n-1}$, it is an easy matter to prove that if $m > n$,

$$|x_m - x_n| = \frac{1}{(n+1)!} + \ldots + \frac{1}{m!} < \frac{1}{2^n} + \ldots + \frac{1}{2^{m-1}} < \frac{1}{2^{n-1}},$$

and so (b) satisfies the Cauchy convergence criterion. Hence if the real numbers are complete there is a limit. This limit is known to be $e - 1$, which is not a rational number.

Now one way to obtain the completeness of our system of real numbers is simply to add this as an additional postulate. This is sometimes done and there is nothing wrong with it. There are several ways of ensuring completeness by the addition of an extra postulate, one of the simplest of which is as follows: let E be a collection of real numbers which has an upper bound; that is, a number M surpassed by no member of E. Then we postulate that among all upper bounds of E there is a smallest one. This least upper bound is frequently denoted by sup E. If we assume this fact about the real numbers, we can prove completeness, and conversely.

Another way to achieve completeness is to construct, using the rational numbers, a more inclusive system which is complete. An extremely direct way to do this would be to define a real number to be a Cauchy-convergent sequence of rational numbers, and to define order and the arithmetic of these sequences. For example, the sum of the sequence $a_1, a_2, \ldots$ and the sequence $b_1, b_2, \ldots$ is the sequence $a_1 + b_1, \ldots$. This definition has the fault that two different sequences which, in the ordinary parlance, have the same limit, would be regarded as different real numbers. Hence what is done is to consider two sequences as equivalent when their difference is a sequence whose n-th term tends to zero, and to define a real number as a class of equivalent sequences. The system thus defined can be shown to be complete and to have all the arithmetic properties which one ordinarily

associates with the real numbers.

There are other, but equivalent, ways of completing the space of rational numbers, the most notable of these being the method of Dedekind sections, for which the reader is referred to the elegant little book *Grundlagen der Analysis* by Edmund Landau.

With the knowledge that the real numbers are complete, one can clear up most of the difficulties which arise in analysis in connection with limiting processes, continuous functions and the like. For example, if one looks back into the material covered in the elementary calculus, he will almost certainly find that no proofs were given for the validity of the comparison test or alternating series test, for the fact that a continuous function on an interval assumes all values between those assumed at the ends, or that a continuous function on a closed interval is bounded and has a largest value, or for many other similar facts. All these are simple and direct consequences of the completeness of the real number system.

Henceforth in this article, to aid the reader's intuition as much as possible, we shall consider the real numbers as being plotted in the usual fashion on a line, and shall speak interchangeably of number and point.

A fundamental fact about the real numbers (or points on a line), and one which is easily proved from their completeness, is the so-called principle of Bolzano-Weierstrass. If E is a set of points on a line (or real numbers) a point y is said to be a limit point of E if given any positive number ϵ, there is a point of E, different from y, whose distance from y is less than ϵ. It follows then that there are infinitely many such points. The principle of Bolzano-Weierstrass states that if E is an infinite bounded point set, it has at least one limit point. (The description ''infinite bounded'' is not contradictory. By ''infinite'' we mean that the number of different points is not finite. By ''bounded'' we mean that all of the different points x satisfy a common inequality, $|x| < M$). This principle will not be proved here, but an example will be given of its use, since this will typify many of the proofs needed to fill up the gaps left in elementary analysis.

Recall that a function f is continuous at x_0 if given $\epsilon > 0$, there exists $\delta > 0$ such that provided $|x - x_0| < \delta$ and provided x is one of the points on which f is defined, then $|f(x) - f(x_0)| < \epsilon$. A function f is defined and continuous at each point of an interval $a \leq x \leq b$ is necessarily bounded thereon. We prove this by contradiction. If it is not so, there is a point x_1 where $|f(x)|$ surpasses 1, there is a point x_2 where $|f(x)|$ surpasses 2, and so on; there is a point x_n where $|f(x)|$ surpasses n. The points x_1, x_2, ..., either are infinite in which case they have a limit point y, or else one is repeated infinitely many times and again may be called y. In either case y plainly belongs to our interval and f is defined and continuous at y. Take 1 to be our

ϵ; then by the definition of continuity there exists an interval about y as center, of width $2\delta > 0$, in which f lies between $f(y) - 1$ and $f(y) + 1$. In this interval f is certainly bounded. But the interval must contain infinitely many of the x_n; hence it contains points where $|f(x)|$ is as large as desired. This is a contradiction and the theorem is proved.

Two other tasks that are undertaken in the theory of functions of a real variable are to set up a system of measuring sets of points and to develop a general theory of integration. These two tasks are closely, in fact inextricably, related, and indeed either may be performed first as a tool for the other. Everyone is familiar with the length, area, and volume of simple geometric figures, but to assign a measure to an arbitrary set of points is easily seen to be a difficult job. For example, the measure of the interval $0 \leq x \leq 1$ is quite satisfactorily taken to be its length, namely, 1; however, what measure shall be assigned to the rational points in this interval? And what to the irrational points? Both these sets appear to be quite thickly distributed in the interval. Yet it would not appear proper to give each measure 1, since one of the cherished ideas of measure is that the whole should be the sum of its parts. How then should the measure be distributed? The following is a sketch of the development of the measure and integration theory originally devised by Lebesgue early in the present century. It is perhaps an old fashioned and unsophisticated program by present standards, but has the advantage of appealing to the geometric intuition.

All point sets considered will be on the line. Any set E of objects whatsoever is said to be denumerable if the members of E can be paired off with the members of the set of all positive integers. The members of a denumerable set E can thus be written $x_1, x_2, x_3, \ldots$, by assigning to each member of E as subscript that integer with which it is paired off. For example, the even integers are denumerable (put $x_n = 2n$, $n = 1, 2, 3, \ldots$), so are the odd integers ($x_n = 2n - 1$), and so are the totality of positive, zero, and negative integers ($x_1 = 0$, $x_2 = 1$, $x_3 = -1$, $x_4 = 1$, $x_5 = -2$, etc.) More surprising, the set of all rational numbers is denumerable! Consider the special case of the rationals between 0 and 1, inclusive. How many can be written as fractions with denominator 1? Just 0/1 and 1/1; call them x_1 and x_2. How many with denominator 2? Just 0/2, 1/2, and 2/2, of which only 1/2 is new; call it x_3. With denominator n there are just $n + 1$ fractions to consider, and not all these are new. It is clear how the counting proceeds. We see then that the rational numbers between 0 and 1 are denumerable, and the case of all rational numbers is not much harder. It is known that the irrational numbers in this interval or in any interval are infinite in number but not denumerable.

Recall that if E is a set of points, a point y is a limit point of E if every interval, however small, about y as center, contains a point of E different from y. A limit point of E may or may not belong to E. For

example the set consisting of 1, 1/2, 1/3, 1/4, ... has 0 as a limit point (the only one) and 0 is not a point of the set E. A set E is said to be closed if every limit point of E is a member of E. A set E is said to be open if every point of E is the center of some interval lying wholly in E. Most sets are neither open nor closed. The set E of points satisfying $0 < x \leq 1$ is not closed since 0 is a limit point of E but not in E; it is not open since 1 is in E but is not the center of any interval lying in E. If we denote by $\widetilde{E}$ the set of points not in E, called the complement of E, it is not difficult to show that if E is closed, $\widetilde{E}$ is open, and vice versa. Incidentally in order to make this last statement make sense without exception, it is necessary to postulate the existence of a special set, called the null or void set without any points at all. It is the complement of the entire line and is both open and closed.

In order to avoid sets of infinite measure, let us consider henceforth only sets belonging to some fixed interval I: $A \leq x \leq B$. By $\widetilde{E}$ we shall mean those points of I not in E. The fundamental fact about open sets is this: any open set Ω is the sum of a finite or denumerable collection of intervals α_i, $i = 1, 2, \ldots$ where each α_i is of the form $a_i < x < b_i$, and the different α_i have no points in common. It is only natural to define the measure $m(\Omega)$ of Ω to be the sum $\Sigma_1^\infty m(\alpha_i) = \Sigma_1^\infty (b_i - a_i)$. (or Σ_1^N in case only N intervals are involved). It is easy to see that this series converges to a sum not exceeding $m(I) = B - A$. Any open set thus has positive measure. If F is a closed set, $\widetilde{F}$ is open (except for a slightly exceptional circumstance at A and B. Any rate $\widetilde{F}$ is a finite or denumerable collection of intervals, two of which may contain an end point or reduce to a single point, and we still define $m(\widetilde{F})$ to be $\Sigma(b_i - a_i)$). We then define $m(F)$ to be $m(I) - m(\widetilde{F})$. Every closed set and every open set in I is now provided with a measure. A single point is a closed set and has measure zero.

Now let E be any set in I. We define the outer measure of E, or $\overline{m}(E)$, to be the greatest lower bound of all $m(\Omega)$, where Ω is an open set containing E. Similarly the inner measure $\underline{m}(E)$ is defined to be the least upper bound of all $m(F)$, where F is a closed set in E. In other words, we enclose E in open sets and see how small the measures of the open sets can be, and insert closed sets in E and see how large their measures can be. It is obvious that $\underline{m}(E) \leq \overline{m}(E)$. A set E for which $\underline{m}(E) = \overline{m}(E)$ is called a measurable set, and the common value of $\underline{m}$ and $\overline{m}$ is written $m(E)$. If E is open or closed, it is measurable and this $m(E)$ reduces to that previously defined.

The remarkable things about measurable sets are these: (1) If E is measurable, so is $\widetilde{E}$. (2) If $E_1, E_2, \ldots,$ is any finite or denumerable collection of measurable sets (with or without common points) then the set ΣE_i of points in at least one E_i is measurable and so is the set

ΠE_i of points common to all E_i. (3) If the E_i have no common points, then $m(\Sigma E_i) = \Sigma m(E_i)$. If one considers how he ordinarily constructs point sets, it becomes apparent how difficult it would be to find a non-measurable point set. As a matter of fact, the very existence of *any* non-measurable point set is a matter debated by mathematicians and logicians to this very day. For more about this, see the last paragraph of this article.

As an example, consider a denumerable set E in I: $x_1, x_2, \ldots$. Given $\epsilon > 0$ construct an open interval a_1, of length $\epsilon/2$ with center at x_1, an open interval a_2 of length $\epsilon/4$ with center at x_2, etc. The set $\sum_1^\infty a_i$ is an open set containing E and of measure not exceeding $\sum_1^\infty \epsilon/2^i = \epsilon$, and so $\overline{m}(E) < \epsilon$ for every $\epsilon > 0$. Thus $\overline{m}(E) = 0$, and since $\underline{m}(E) \leq \overline{m}(E)$, we have $m(E) = 0$. In answer to a question previously raised, we see that the rational points in I have measure zero and the irrational points have all the measure.

One might raise the following question: since there is a possibility that non-measurable sets exist but every set has an outer measure, why not use this outer measure as actual measure? The answer is simple. Even if E_1 and E_2 are without common points, one cannot prove that $m(E_1 + E_2) = \overline{m}(E_1) + \overline{m}(E_2)$. The best one can do is to show that $\overline{m}(E_1 + E_2) \leq \overline{m}(E_1) + \overline{m}(E_2)$.

Now let us turn to integration. Let $f(x)$ be a bounded function defined on an interval I, say, $0 \leq x \leq 1$. Recall that the Riemann integral (the ordinary integral of elementary calculus) of f is defined to be the limit of $\Sigma f(x_i)\Delta x_i$ in case this limit exists, the limit here being taken as the maximum of the Δx_i tends to zero. It is an easy matter to give an example of a function whose Riemann integral does not exist; in fact the function which is 1 if x is rational and 0 if x is irrational has no Riemann integral. This function may seem to be an extremely artificial one, but it is the limit of a sequence of Riemann integrable functions. If $f_n(x)$ is defined to be zero at all points except at the first n rational points (arranged in denumerable order), at which points $f_n(x) = 1$, we see that $\lim f_n(x) = f(x)$. Now why was this fact mentioned? Because frequently in analysis a function is arrived at as the limit of other functions, and nothing more may be known about it. Even though we may integrate the approximating functions, we may not so integrate the limit function, and the frequently desirable property $\lim \int f_n(x)dx = \int \lim f_n(x)dx$ does not hold. This gives a suggestion of the desirability of a more widely applicable integration process.

Let us consider then a function f defined on a measurable set E, which may, of course, be an interval. f is said to be a measurable function if for every number a, the set of points of E for which $f(x) > a$ is a measurable set. In this definition the sign $>$ may be replaced by

$\geq$, $<$, or $\leq$; all are equivalent. It is clear that finding a non-measurable function is exactly the same problem as finding a non-measurable set. One can show that all the arithmetic operations on measurable functions yield measurable functions, and that the limit of measurable functions is a measurable function.

Now suppose that f is bounded, $m \leq f < M$ on E. Insert numbers l_1, l_2, ... , l_k between m and M such that $m = l_1 < l_2 < \ldots < l_k = M$. Let E_i be the set on which $l_i \leq f < l_{i+1}$, and consider the sum $\Sigma l_i m(E_i)$. With no further hypotheses on f, it can be shown that as the maximum $l_{i+1} - l_i$ tends to zero, this sum has a limit which is called the Lebesque integral of f over the set E, and which we shall denote by ${}^L\int_E f(x)dx$. If E is an interval and if the Riemann integral of f exists, the two integrals are the same. In the case of the function f which is zero on the irrationals and one on the rationals, the Lebesgue integral is sero, since the function is zero except on a set of measure zero.

In order to see better how this integral is defined and how it differs from the Riemann definition, consider as a specific example the function $y = f(x) = |x|$ on $-1 \leq x \leq 1$. Divide the positive y axis by values l_1, l_2, etc. spaced at intervals of 1/10, starting at $y = 0$. The set E_1 is the interval $-1/10 < x < 1/10$, E_2 consists of the two intervals $-2/10 < x \leq -1/10$ and $1/10 \leq x < 2/10$, etc. The sum $\Sigma l_i m(E_i)$ is easily recognizable as the area of a stair-step figure, rising in both directions from the origin and remaining under the graph of f. For such a simple function as this, of course, there is not much difference between the Riemann and Lebesgue methods. However, for a complicated function, the Lebesgue method can create a subdivision of I into sets which are altogether different from anything allowable in the Riemann process.

If f is unbounded, what we do is to cut off the graph of $f(x)$ with horizontal lines at height $+N$ and $-M$ to obtain a bounded function which we may call $f_M^N(x)$. This bounded function is then integrated and N and M allowed to become infinite independently. If the limit of the integral of f_M^N exists, it is called ${}^L\int_E f(x)dx$ and f is said to be Lebesgue summable on E. As a result of this definition, f is summable if and only if $|f|$ is summable. The corresponding fact is not true in Riemann integration and as a result, the function $x^{-1}\sin x^{-1}$ is not Lebesgue summable in $0 \leq x \leq 1$, but is Riemann (improperly) integrable.

The Lebesgue integral thus defined has many useful properties. For continuous functions it is the same as the Riemann integral. For bounded functions at least, it exists whenever f is measurable, and this allows a great degree of discontinuity in f. It is too much to hope that for a sequence $f_n(x)$ of integrable functions, we have ${}^L\int_E \lim f_n(x)dx = \lim {}^L\int_E f_n(x)dx$, but this is true in a number of important special cases, one of which is the case in which all $|f_n(x)|$ have a common bound.

It might be well to mention here an important application of Lebesgue integration, that to the theory of Fourier series. For this theory the class of functions whose squares are Lebesgue summable plays a special role. Call this class L_2. (As an example, consider $f(x) = x^{-\frac{1}{2}}$. It is Lebesgue summable on $0 \leq x \leq 1$ but its square is not; hance f is not in L_2). It is not difficult to show that if f is in L_2 and if we denote the Fourier coefficients of f by a_n and b_n, then $\pi\sum_0^\infty(a_n^2 + b_n^2)$ converges; in fact the sum does not exceed $^L\int_0^{2\pi} f^2(x)dx$. This fact (Bessel's inequality) will be found in some form or other in most elementary books on Fourier series. However, what is more remarkable is this: If a_n and b_n are any sequences of numbers such that $\sum_0^\infty(a_n^2 + b_n^2)$ converges, there exists a function in L_2 whose Fourier coefficients are precisely the a_n and b_n. This function is unique in the sense that if two such functions exist, they can differ only on a set of measure zero. This is the famous theorem of Riesz-Fischer. If only Riemann integration is allowed, the theorem is no longer true.

The fundamental theorem of calculus also appears in the Lebesgue theory. Let $f(x)$ be summable on an interval I and let a and x be in I. Then $^L\int_a^x f(t)dt$ is a function of x, say $F(x)$. The derivative $F'(x)$, computed by the usual Δ-process, need not exist for all x; however, it does exist and equal $f(x)$ at all points of I with the exception of a set of measure zero.

The Lebesgue theory of measure and integration can be extended and generalized in various directions. One of these is its extension to the plane, three-space, or even higher dimension spaces. In the plane, for example, the rectangles play the fundamental role that intervals play on the line. Another generalization is this: Consider a function $\phi(x)$ defined on some interval I, such that $\phi(x)$ always increases or remains constant when x increases. For any sub-interval α: $a \leq x \leq b$ of I, let the measure of α be $\phi(b) - \phi(a)$ instead of merely $b - a$. A great deal of the measure and integration theory can be developed for this situation. The resulting integral is called the Lebesgue-Stieltjes integral. It arises, for example, in physical problems in which we have a distribution of positive mass along a line and wish to integrate with respect to the mass, as in the computation of a moment of inertia, $\int x^2 dm$.

The reader will find, if he looks into books on the subject, approaches to the Lebesgue theory which look quite different from that just outlined. There is a most elegant approach, developed by F. Riesz, in which sets of zero measure are first defined, and then a function is defined to be measurable if it is the limit, almost anywhere, of a sequence of step functions. (A step function has a graph consisting of a succession of horizontal segments; almost everywhere means with

the exception of a set of measure zero). The integral is then defined, and the measure of a set E is taken to be the integral of a function which is 1 on E and 0 off E.

We have attempted to outline some of the central ideas, methods, and facts that appear in the theory of functions of a real variable. The theory is extensive, and has many ramifications and applications not even hinted at here. It is a fairly severe discipline. The books and papers on the subject do not read like novels, but usually require considerable application. However it is a useful, fascinating, and it is believed by the author of this article, a rewarding discipline.

In closing, it may entertain the reader to include an example of a non-measurable set. In order to avoid prolixity, let it be supposed the Lebesgue measure has been developed for the circumference C of a circle of perimeter π instead of for a line. This should not put a great strain on the imagination. If P is a point on C, we consider the set E_P of points which can be reached from P by stepping of distances 1, 2, 3, ... along C in both directions. E_P is denumerable, and if Q is in E_P, then $E_P = E_Q$. Consider the totality of all such sets E_P obtained in this way, and let L_0 be a set consisting of exactly one point from each E_P. Let L_n be the set obtained by rotating L_0 in the counterclockwise direction by n units, and L_{-n} the set similarly obtained by rotating L_0 in the clockwise direction. All the L's are congruent, no two have a point in common, and together they make up C. If L_0 is measurable, so are the other L's and all have the same measure. Then $\pi = m(C) = \Sigma_{-\infty}^{\infty} m(L_n) = \Sigma_{-\infty}^{\infty} m(L_0)$. This is plainly impossible, since this series is zero if $m(L_0) = 0$ and diverges if $m(L_0) \neq 0$. Hence L_0 is non-measurable. In this construction the selection of the set L_0 so as to have exactly one point from each E_P is the debatable step. It involves the use of a famous axiom of logic, called the axiom of choice, the permissibility of which is in dispute among logicians.

University of California at Los Angeles

CURRENT PAPERS AND BOOKS

Edited by

H. V. Craig

This department will present comments on papers previously published in the MATHEMATICS MAGAZINE, lists of new books, and book reviews.

In order that errors may be corrected, results extended, and interesting aspects further illuminated, comments on published papers in all departments are invited.

Communications intended for this department should be sent in duplicate to H. V. Craig, Department of Applied Mathematics, University of Texas, Austin 12, Texas.

Calculus and Analytic Geometry. By Cecil Thomas Holmes. New York, McGraw-Hill Book Company, Inc., 1950. 10 + 416 pages.

This text offers numerous features which are attractive. The first chapter of the analytic geometry begins with introductory theory of equations. In the initial pages the graphical method of approximating the real roots of an algebraic equation is given; this process is well placed in view of the occurrence of higher degree equations later in analytic geometry and the calculus. In many calculus books the higher degree equations are carefully avoided.

The usual material of an analytic geometry text through the straight line appears in the first chapter. The establishing of the equations of loci and plotting loci are found in this chapter. The calculus begins in chapter 2. The student is taken into the idea of the derivative as applied to the polynomial after little preliminaries on limits and continuity. This chapter ends with simple rate problems.

The analytic geometry is woven into the calculus throughout the text. Chapter 3 consists of material on the conic sections. Chapter 4 is a brief treatment on maxima and minima problems. Chapter 3 prepares the student to handle certain problems involving the conics in chapter 4. This is effective pedagogy.

In Chapter 5 the student meets the idea of the integral which deals only with the polynomial integrand. In this chapter is found methods for finding volumes of solids of revolution, volumes of solids of known cross section, areas between curves, pressure on a vertical wall, the work done in pumping a liquid from a tank, and the usual elementary methods of approximating an integral. The treatment of the applications at this point seems to be too ambitious for students so immature in thinking the calculus.

The deferred treatment of the product and quotients rules for the derivative until chapter 6 is noteworthy. These rules are applied in this chapter to algebraic functions, and immediately following these exercises the integrals involving algebraic functions appear. Again, this appears to be a sound teaching gesture. Chapter 7 deals with the

arc length, surfaces of revolution, parametric equations, and curvilinear motion. Ending with chapter 7 the text offers a course in the analytic geometry and the elementary calculus of algebraic functions; this is a well-arranged unit.

Chapters 8 to 10, inclusive, treat the usual differential and integral calculus of the simple transcendental functions. Chapter 9 gives the derivations of the formulas from trigonometry used in the calculus. In this chapter loci in polar coordinates are taken up. In chapter 11 infinite series and expansions of functions are found. Twenty-six pages on ordinary differential equations make up the material of chapter 12. This occurrence of the differential equations prior to the multiple integrals is unusual. Chapter 13 contains solid analytical geometry preparatory to the calculus of several variables. Chapters 14 and 15 complete the text with partial derivatives and the multiple integrals.

Several miscellaneous observations may be pertinent. More illustrative examples in places would help. Examples of this need may be noted in treating curve sketching in Section 15 and the derivative from its definition in Sections 20, 21, and 22. The notation "$D_x Y$" for the derivative prior to the use of the differential of a function is quite desirable. It is notable and desirable, I think, to find given on page 111 the general form of the fundamental theorem of the integral calculus. The irregularity in giving answers will not appeal to many instructors. Usually the even or odd numbered or all answers are given. Since the text covers analytics and the calculus, a collection of numerical tables would have been worthwhile.

The arrangement of topics in the text is unlike the usual calculus text but seems to be logical and pedagogically sound. A random sampling revealed no incorrect answers. The figures are clear and pleasing and the format is quite good.

P. K. Smith

High-Speed Computing Devices. By the staff of Engineering Research Associates, Inc., New York (McGraw-Hill Book Company) 1950; 450 pages $6.50.

Many of the applications of mathematics are now being exploited because of the rapid development (brought on by World War II) of high-speed computing devices. Even pure mathematics stands to gain much from the availability of this new equipment. In this connection the fine tribute paid by G. H. Hardy to the computational work of Major MacMahon might be noted:

"At this point we might have stopped had it not been for Major MacMahon's love of calculation. MacMahon was a practiced and enthusiastic computer, and made us a table of $p(n)$ up to $n = 200$. In particular he found that

$$p(200) = 3,972,999,029,388$$

> and we naturally took this value as a test for our asymptotic formula. We expected a good result, with an error of perhaps one or two figures, but we had never dared to hope for such a result as we found. Actually 8 terms of our formula gave $p(200)$ with an error of 0.004. We were inevitably led to ask whether the formula could not be used to calculate $p(n)$ *exactly* for any large n." [1]

If Prof. Hardy could make such a statement stressing the importance of numerical work, who is there to say that calculation plays no part in the development of pure mathematics?

The book under review contains many electrical circuit diagrams and considerable discussion on the physical components of the computing devices. These components have been referred to as the "hardware" of the system by Prof. D. R. Hartree. Inasmuch as these parts of the book are essentially non-mathematical in character, they will not be dealt with in this review. Of much more mathematical importance is the section on machine commands, which gives a very excellent description of how the "hardware" receives instructions and translates them into concrete numerical steps. Indeed, the choice of these commands requires that the intended mathematical user and the constructing engineer combine their talents and each learn a little of the other's vocabulary and problems.

There have been several recent books which have approached this meeting of the minds, ranging from a popular treatment to some very specialized treatises on particular phases of this problem. But the volume under review appears to be the most suitable yet published for the use of research workers and graduate students who desire to learn something of this new field.

Of the two purely mathematical chapters, one is on arithmetic systems, and seems to be in good mathematical order. The second is on numerical analysis and its contents have been admittedly borrowed from some of the standard works on that subject. The reader is cautioned to go to the original sources here, for this chapter contains a number of errors. For example on page 112 in discussing Weddle's rule it is stated that all differences above the sixth are neglected and hence implied that Weddle's rule is correct through sixth differences. This is not quite the case. On page 113 it is stated that the Gauss method of numerical integration gives an *exact* result where "the order of the function is less than or equals $2n - 1$." Presumably what is meant here is "the function is a polynomial of degree less than or equal to $2n - 1$." The notation on page 124 explaining Graeffe's method is unsuitable, it certainly does not follow that "$f_2(x) = f_1(x)f_1(-x)$" as is used in the text. The use of "$f_2(x^4) = f_1(x^2)f_1(-x^2)$" would be a considerable improvement here. At other places in this chapter, "root" is used where "zero" would be more appropriate, and also "root" is used where clearly "approximate root" is meant. Indeed, this chapter on numerical analysis bears the stamp of an enthusiastic physicist rather than the stamp of a careful

mathematician.

Despite this adverse criticism, the book is quite suitable for use by graduate students, most of whom will recognize the errors noted above and supply new and correct terminology as they proceed. A valuable feature is the bibliography at the end of each chapter, which gives numerous references to original papers and articles dealing with the development of particular devices and methods.

Finally, the reviewer would like to add some comment of his own on the growth of techniques applicable to high-speed computers. Lord Kelvin once stated:

> "When you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."

Most scientific men now agree that this dictum of Lord Kelvin is unworthy of a present day scientist. It is perhaps, therefore, of some interest to note that although computers deal essentially with a numbering system, there are many so-called logical problems capable of being "programmed" or set up for solution by the use of commands in a high-speed computer. Dr. Shannon of the Bell Telephone Laboratories has recently described the possibility of programming a machine for chess playing.[2] Machines have been designed to play the game of Nim, and this reviewer has heard of a machine, which when finding a human opponent who makes the correct move in Nim three times in succession and thus gives the machine a losing combination three times in succession, then resigns by operating a typewriter which spells out "RESIGN".

The mathematical approach needed for formulating programs for numerical problems (or for logical problems) is stressed in the volume under review, and if this volume serves to stimulate programming among mathematicians, it will have rendered a distinct service.

1. G. H. Hardy, *Ramanujan* (Cambridge) 1940, p. 119. The reviewer is indebted to Prof. Hans Rademacher for supplying this reference.

2. Claude Shannon, "Programming A Computer for Playing Chess", *Philosophical Magazine* series 7, v. 41, pp. 256-275 (1950).

Robert E. Greenwood

THE PERSONAL SIDE OF MATHEMATICS

WHAT MATHEMATICS MEANS TO ME

Aristotle D. Michal

As I look back, the mathematics to which I contribute and like best has been colored by pondering over certain statements or claims made by my teachers at critical junctures in my high school and graduate student years. My first resolve to aim at a professional mathematical career was made as a result of a very laudatory remark made in 1915 by H. B. Marsh, my Springfield, Mass. high school teacher, in a review course in geometry.

When I was looking around in the spring of 1921 for a place to continue my graduate work, my Clark University teachers Taber and Webster gave me wise spontaneous counsel that still "rings in my ear"! Webster informed me of a Volterra who preferred working with analytic functions of an infinite number of variables and of an Evans who was a leader in the most recent work on integral equations and functionals. Webster in an inspired semi-popular talk on general relativity - given during Einstein's first visit in the U.S.A. - spoke eloquently of the wonders of the tensor calculus and differential geometry. Taber's casual remark that one theorem in Lie's continuous groups explains the integration tricks in elementary differential equation theory, fired my imagination.

Mathematics has a fascination that is often brought about by the realization (often like a flash) in one's work that a certain subject habitually considered far removed from one's endeavors, suddenly takes on the status of a close neighbor. Such was the case in my work of the early and middle twenties on continuous groups and semi-groups in function spaces and infinite-dimensional differential geometry on one hand, and abstract spaces and abstract algebra on the other. To take a more specific example, I suddenly realized in 1931 that the development of abstract normed linear rings (with or without involution) would be very useful in studying the generalization of the Ricci tensor (Einstein tensor) in infinitely-dimensional Riemannian differential geometries.

What is now needed is a far reaching "*geometrization of analysis*" with the aid of the concepts of *differential geometry*. So far, only a thorough "elementary geometrization" has taken place since Hilbert space and normed linear spaces are generalizations of Euclidean and other elementary spaces.

My passion for mathematics has many facets. Mathematical activity gives me the opportunity to engage in a great sport from both the pure and the applied angles. But mathematics for me is also a *fine* art and in so considering the subject, it contributes to the satisfaction of my longings for the "good and the beautiful".

MISCELLANEOUS NOTES

Edited by

Charles K. Robbins

Material for this department should be sent to C. K. Robbins, Department of Mathematics, Purdue University, Lafayette, Indiana.

This letter is in reference to a letter in Mathematical Miscellany of the Nov.-Dec. 1950 issue of Mathematics Magazine on solving for the length of the hypotenuse of a right triangle on a slide rule.

A very few electrical engineers use the following method for obtaining the length of the hypotenuse. Let us assume that the lengths of the two sides are 18 and 28.

1. Place the left unity of the *C* scale over 18 on the *D* scale.
2. Place the hairline (cursor) over 28 on the *D* scale and read 2.42 on the *B* scale.
3. Add 1 to 2.42 to obtain 3.42 and place the hairline over 3.42 on the *B* scale. Read the answer 33.3 on the *D* scale.

Only one setting of the slide and two of the hairline are required, so that this process is as fast as any for a slide rule. Since in most calculations the electrical engineer is interested in the angles in the triangle, he uses one of the other methods using the trigonometric scales. On the log log duplex vector slide rule this requires two settings of the slide and one of the hairline. On the log log duplex decitrig slide rule one setting of the slide and two of the hairline are required. Both of these methods give the angles.

H. F. Mathis

The problem you propose on p. 115 of the Nov.-Dec. 1950 issue may be more difficult than the solution of a non-linear differential equation, but here is my try:

I present the explanation in skeletal form. The missing words are obvious:

1. We have $\left(\dfrac{28}{18}\right)^2 \approx 2.42$

2. Since $\left(\dfrac{18}{18}\right)^2$ or 1^2 or $1 = 1$

3. We have, by addition, $\left(\dfrac{28}{18}\right)^2 + \left(\dfrac{18}{18}\right)^2 = \dfrac{28^2 + 18^2}{18^2} = 3.42$

4. Therefore $\dfrac{\sqrt{28^2 + 18^2}}{18} = \sqrt{3.42}$ (1.85)

5. Therefore $\dfrac{\sqrt{28^2 + 18^2}}{18} \cdot 18$ or $\sqrt{28^2 + 18^2} = 1.85 \times 18$ (33.3)

Charles Salkind

PROBLEMS AND QUESTIONS

Edited by

C. W. Trigg, Los Angeles City College

Readers of this department are invited to submit for solution problems believed to be new and subject-matter questions that may arise in study, in research, or in extra-academic situations. Proposals should be accompanied by solutions, when available, and by such information as will assist the editor. Ordinarily, problems in well-known text-books should not be submitted.

In order to facilitate their consideration, solutions should be submitted on separate, signed sheets within three months after publication of the problems. Readers are invited to offer heuristic discussions in addition to formal solutions. Manuscripts typewritten on 8½" by 11" paper, double-spaced and with margins at least one inch wide are preferred. Figures should be drawn in india ink and twice the desired size.

Send all communications for this department to *C. W. Trigg, Los Angeles City College, 855 N. Vermont Ave., Los Angeles 29, Calif.*

PROPOSALS

91. *Proposed by Cyril Tobin, St. Francis Xavier University, Antigonish, Nova Scotia.*

An old man had three married daughters, Mrs. Jones, Mrs. Smith, and Mrs. White. Each daughter had one and only one child. When the old man died, following the terms of his will, his fortune was distributed among his daughters, sons-in-law, and grandchildren in the following manner.

Each of the nine persons received a number of envelopes, (the number of envelopes being different for each person). In each envelope was found a number of dollars equal to the number of envelopes which the person received. The share of money for each woman exceeded that of her husband by the amount that her husband's share exceeded that of their child. Although no two families received the same amount of money, all three families did receive the same number of envelopes. The total number of envelopes received by Mrs. Jones and Mrs. Smith was equal to the total number of envelopes received by Mrs. White and Mr. Jones.

Before the will was read all nine persons were penniless. After its administration, all were wealthy but not one of them was a millionaire. What were the nine sums of money?

92. *Proposed by P. A. Piza, San Juan, Puerto Rico.*

Prove that $2n + 1$ is a prime if $\binom{n+r}{2r}/(2r+1)$ is an integer for $r = 1, 2, 3, \dots, (n-1)$, and conversely.

93. *Proposed by Leo Moser, Texas Technological College.*

What is the size of the largest square that can be completely covered by three unit squares?

94. *Proposed by Michael Goldberg, Washington, D.C.*

Find the largest regular octahedron which can be inscribed in a cube.

95. *Proposed by A. O. Qualley, Farnhamville Independent School, Iowa.*

Each side of a square sheet of paper $ABCD$ equals $2a$. The midpoints of AB, BC, CD, and DA are E, F, G, and H, respectively, and O is the center of the square. The square is creased inward along EG and FH, and outward along the diagonals AC and BD. E, F, G, and H are then drawn inward to form the base of a square pyramid O-$EFGH$, surrounded by four tetrahedra $OAHE$, $OBEF$, $OCFG$, and $ODGH$. (1) If each side of the base $EFGH$ is a, find the volume of the whole solid. (2) Find the length of the side of the base $EFGH$ for which the total volume is a maximum, and compute the maximum volume.

96. *Proposed by R. E. Horton, Los Angeles City College.*

Determine conditions on the integers p and a such that both coordinates of the points of intersection of $x^2 + y^2 = a^2$ and $y^2 = 2px$ will be integers.

97. *Proposed by Bruce Kellogg, Massachusetts Institute of Technology.*

Let $\{i_n\} = i_1,\ i_2\ \ i_3, \cdots$ be a sequence of real numbers such that $\lim\limits_{n\to\infty} i_n = 1$ and $i_n > 1$ for all n. Does the infinite series $\sum\limits_{n=1}^{\infty} 1/n^{i_n}$ converge or diverge, or does the divergence or convergence depend on the sequence $\{i_n\}$?

SOLUTIONS

Late Solutions

64. *P. N. Nagara, College of Agriculture, Bangkok, Thailand.*

Maximum Evaporation Surface on a Spinning Disk

65. [May 1950] *Proposed by Howard Eves, Oregon State College.*

A thin disk of a given radius is spinning on a fixed horizontal axis, and while spinning dips into a vat of liquid. Locate the position of the axis with respect to the surface of the liquid so that the spinning disk exhibits a maximum evaporation surface.

Solution by W. R. Talbot, Jefferson City, Missouri. The wet portion on each side of the stationary disk will be a segment (S) of a circle, and the spinning of the disk will give an annulus (T). If r is the radius of the disk and x is the distance from the axis to the liquid surface, then the area of the evaporation surface is

$$A = T - S = \pi(r^2 - x^2) - [r^2 \text{ arccos } (x/r) - x\sqrt{r^2 - x^2}\,] .$$

Setting $dA/dx = 0$ gives $-2\pi x + 2\sqrt{r^2 - x^2} = 0$, from which

$$x = r/\sqrt{1 + \pi^2}.$$

Also solved by Dewey Duncan, East Los Angeles Junior College; E. S. Keeping, University of Alberta; and the proposer, who remarked that the problem arose in connection with the commercial concentration of fruit juices.

Two Circles and an Ellipse.

66. [May 1950] *Proposed by E. P. Starke, Rutgers University.*

Given four concyclic points A, B, O, P such that $AO = BO$. Show that the ellipse through P having foci at A and B, and the circle through P having center at O are orthogonal or tangent according as O, P are on the same side of AB or on opposite sides.

Solution by R. E. Horton, Los Angeles City College. Suppose first that O, P are on the same side of AB. Now the tangent RPS to the ellipse makes equal angles with the lines PA, PB to the foci, that is, $\angle SPB = \angle RPA$. If BP be extended to C, then $\angle SPB + \angle RPA = \angle CPR + \angle RPA = \angle PAB + \angle PBA = \frac{1}{2}$ arc AOB. It follows that $\angle RPA = \frac{1}{4}$ arc AOB. But $\angle OPA = \frac{1}{2}$ arc $AO = \frac{1}{4}$ arc AOB. Therefore RP coincides with OP and hence is perpendicular to the tangent to the circle at P, so the curves are orthogonal.

Now suppose that O, P are on opposite sides of AB, with RPS tangent to the ellipse at P. Again, $\angle SPB = \angle RPA$. Since $BO = AO$, $\angle BPO = \angle APO$. Adding, $\angle SPO = \angle RPO = 90°$, since these angles are supplementary. Now since RPS is perpendicular to OP, RPS coincides with the tangent to the circle at P, so the curves are tangent.

Also solved by W. B. Clarke, San Jose, California; Howard Eves, Oregon State College; P. N. Nagara, College of Agriculture, Bangkok, Thailand; Gerhard Rayna, New York, N.Y.; P. D. Thomas, Washington, D.C.; and the proposer.

Two Orthogonal Spheres

67. [May 1950] *Proposed by N. A. Court, University of Oklahoma.*

Given two spheres (A), (B). Let p be the perpendicular dropped from B upon a plane (Q) passing through A. If the traces of p and (Q) upon the plane of the common circle (S), real or imaginary, of the two spheres are pole and polar with respect to (S), the two spheres are orthogonal.

Solution by P. D. Thomas, U. S. Coast and Geodetic Survey, Washington, D.C. Let the radius of the common circle (S) be $\pm r$. A plane (Q) through A meets the plane of (S) in the line l. The perpendicular p from B upon (Q) meets the plane of (S) in the point P. From this construction the plane $ASBP$ is necessarily perpendicular to l at a point say O, and from the similar right triangles AOS and PBS we have

$$AS \cdot SB = SO \cdot SP. \tag{1}$$

But if P and l are pole and polar with respect to (S), then

$$SO \cdot SP = r^2,$$

whence from (1),

$$AS \cdot SB = r^2,$$

which is the condition that the spheres (A) and (B) should be orthogonal.

Also solved by A. Sisk, Maryville College, Tennessee.

The Pineapple Purchasers

70. [Sept. 1950] *Proposed by J. S. Cromelin, Clearing Industrial District, Chicago, Illinois.*

Smith, Brown, Jones, and Robinson were married to Minnie, Maggie, Maud, and Mabel, but not in that order. (1) Each one bought as many pineapples as he or she paid cents apiece for them. (2) Mabel bought four times as many as Mr. Jones, and (3) Mr. Brown paid three cents apiece more than his sister, Maud. (4) Each wife spent \$1.05 less than her husband, and (5) Mrs. Robinson spent more than Mrs. Smith. From these data the full names of the two ladies visiting Hawaii at the time may be determined. Who were they?

Solution by P. N. Nagara, College of Agriculture, Thailand. If the husband paid h cents apiece and the wife paid w cents apiece, then from (1) and (4),

$$(h - w)(h + w) = h^2 - w^2 = 105 = 1 \cdot 105 = 3 \cdot 35 = 5 \cdot 21 = 7 \cdot 15.$$

Clearly, the only integer solutions (h,w) of this equation are (53,52), (19,16), (13,8), and (11,4). From (2), Jones bought 13 and Mabel bought 52. From (3), Brown paid 11 cents apiece and Maud paid 8 cents, so her full name was Mrs. Maud Jones. From the data, Minnie and Maggie cannot be completely identified, hence from (5) the other Hawaiian visitor was Mrs. Mabel Robinson.

Also solved by *R. P. Banaugh, El Cerrito, California; Monte Dernham, San Francisco, California; Fred Marer, Los Angeles City College;* and *F. L. Miksa, Aurora, Illinois.*

Dernham claims that the problem data are insufficient for a unique solution and that the solution above implies a sixth condition, "that each of the eight persons bought a different number of pineapples". Without this restriction, Brown may have bought 19 and his sister Maud 16, provided Brown's wife also bought 16 and Maud's husband 19. In this event Maud would have been Mrs. Maud Smith, while the other visitor was still Mrs. Mabel Robinson.

A Sixth Power Digit Puzzle

71. [Sept. 1950] *Proposed by Victor Thebault, Tennie, Sarthe, France.*

Using each of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 once and only once, form three numbers, m, n, r, such that the sixth power of r equals the sum of m and n.

Solution by Fred Marer, Los Angeles City College. $r^6 = m + n$, and since the ten and only the ten digits are to be used once each, m may not contain more than seven digits. So $4 < r < 15$. Now any sixth power is congruent to 0 or to 1 modulo 9, so $m + n \equiv 0$ or $1 \pmod 9$. Since the sum of the ten digits is 45, it follows that r is either 8 or 9. Hence n has three digits. Brief experimentation reveals that there are three basic solutions:

$$261790 + 354 = 8^6, \quad 261709 + 435 = 8^6, \text{ and } 530827 + 614 = 9^6.$$

In each solution, the hundred's digits of m and n may be interchanged without changing the value of $m + n$. The same is true for the ten's digits and for the unit's digits. Also, m and n are interchangeable. Thus the problem has $2^3 \cdot 3 \cdot 2$ or 48 solutions.

Also solved by *Monte Dernham, San Francisco, Calif.; P. N. Nagara, College of Agriculture, Thailand; F. W. Saunders, Coker College, Hartsville, S. C.; E. D. Schell, Arlington, Va.; and the proposer.* One incomplete solution was received.

A Triangle Construction

72. [Sept. 1950] *Proposed by D. L. Mackay, Manchester Depot, Vermont.*

Construct a triangle given, in position, a vertex A, the foot D of the corresponding altitude, and a point M which divides a second altitude in a given ratio $p:q$.

Solution by A. L. Epstein, Waltham, Mass. Draw line a perpendicular to AD at D. Through M draw an arbitrary line meeting a in P. Locate Q so that $PM:MQ::p:q$. Draw line f through Q and parallel to a. On AM as a diameter describe a circle intersecting f in E and E'. Draw EM meeting a in B and AE meeting a in C. Then ABC is the required triangle. A second triangle is obtained from E'. There will be but one solution if M falls on AD, or no solution if A lies between a and f.

Also solved by *Howard Eves, Oregon State College; H. E. Fettis, Dayton, Ohio; Frank Hawthorne, Hofstra College; L. M. Kelly, Michigan State College; P. N. Nagara, College of Agriculture, Thailand; C. C. Richtmeyer, Central Michigan College of Education; and the proposer.*

A Trigonometric Equation

73. [Sept. 1950] *Proposed by W. E. Byrne, Virginia Military Institute.*

Consider the real roots in the interval $-\pi \leqq \phi \leqq \pi$ of the equation

$$f(\phi) = \omega^2 b \sin 2\phi - 3(g \sin\phi - \omega^2 a \cos\phi) = 0,$$

where ω, a, b, g are positive constants. Show that there may be two or four distinct roots, or three distinct roots of which one is double.

What relations must hold among the constants to give a double root?

I. *Solution by H. E. Fettis, Dayton, Ohio.* With $3a/2b = m$, $3g/2b\omega^2 = n$, $\sin\phi = x$, the given equation may be written

$$\pm\sqrt{1 - x^2}\,(x + m) = nx, \tag{1}$$

or, upon squaring and expanding,

$$F(x) = (x^2 - 1)(x + m)^2 + n^2x^2 = 0. \tag{2}$$

Since

$$F(-1) = n^2, \quad F(0) = -m^2, \quad F(1) = n^2,$$

there is at least one root in the interval $0 < x < 1$ and one in the interval $-1 < x < 0$. Also, by Descartes' rule, there is only one positive root.

The resolvent cubic of (2) is

$$y^3 - (m^2 + n^2 - 1)y^2 - 4m^2n^2 = 0.$$

The discriminant of this cubic is, except for a positive factor, the discriminant of (2). It is

$$12m^2n^2\,[27m^2n^2 + (m^2 + n^2 - 1)^3]. \tag{3}$$

If (3) is negative, the roots of (2) are all real or all imaginary (but the possibility of all roots imaginary has been eliminated above.) If (3) is positive, there will be only two real roots. If (3) vanishes, (2) has a double root. Now, by hypothesis, $m^2n^2 > 0$. The last factor of (3) may be resolved into

$$m^{2/3} + n^{2/3} - 1$$

multiplied by other factors which are positive for all real values of m and n. Thus the equation (2) has four real and distinct roots, three distinct roots of which one is double, or only two real roots according as

$$m^{2/3} + n^{2/3} - 1 \lesseqgtr 0.$$

When there is a double root, it is easily found to be $x = -m^{1/3}$.

From (2) it is evident that $F(x) = 0$ can have no root whose absolute value exceeds 1, i.e. all real roots of (2) give real values of ϕ. Finally we must choose that solution of $\phi = \arcsin x$ in the interval $-\pi < \phi < \pi$ for which

$$\cos\phi = \pm\sqrt{1 - x^2}$$

has the sign determined by (1). Thus, if x is positive, $\cos\phi$ must be positive; while if x is negative, $\cos\phi$ is positive or negative according as $|x| \gtrless m$.

II. *Solution by the Proposer.* The original equation reduces to

$$n\tan\phi = \sin\phi + m.$$

The graphs of $y = n \tan \phi$ and $y = \sin \phi + m$ evidently intersect just once in each of the intervals $-\pi < \phi < -\pi/2$ and $0 < \phi < \pi/2$. In the interval $-\pi/2 < \phi < 0$ they are tangent provided y and $dy/d\phi$ have the same values for both curves at the same point. This occurs if

$$m^{2/3} + n^{2/3} = 1, \quad \sin^3\phi = -m, \quad \cos^3\phi = n.$$

It is then easy to show there are two intersections in this interval, or none, according as

$$m^{2/3} + n^{2/3} \lessgtr 0.$$

The problem arose in connection with the discussion of the stationary conditions (stable or unstable equilibrium) of a homogeneous bar *BC* of mass *m*, attached to a bar *AB*. *AB* rotates in a horizontal plane with constant angular speed ω about the fixed point *A*. *BC* is hinged so as to remain in the vertical plane through *AB*.

Also solved by *P. N. Nagara, College of Agriculture, Thailand.*

QUICKIES

From time to time this department will publish problems which may be solved by laborious methods, but which with the proper insight may be disposed of with dispatch. Readers are urged to submit their favorite problems of this type, together with the elegant solution and the source, if known.

Q 28. Simplify $(27 + 8i)/(3 + 2i^3)$.

Q 29. Given $y_1 = x$, $y_n = y_{n-1}^x$. Find dy/dx. [*Submitted by Leo Moser.*]

Q 30. Prove Proposition XII, Second Book of Apollonius, Conics: If from a point on a hyperbola lines are drawn in two given directions to meet the asymptotes, the product of the two distances is independent of the position of the point. [*Submitted by Fred Marer.*]

Q 31. Two fixed lines are perpendicular to each other. An ellipse moves so that it is always tangent to both lines. Find the locus of its center. [*Submitted by Adrian Struyk.*]

Q 32. Each edge of a cube is a one-ohm resistor. What is the resistance between two diagonally opposite vertices of the cube? [*Submitted by Nathan Eisen.*]

Q 33. Prove that the sum of the divisors of a number, divided by the number itself, is in general larger than the sum of the squares of the divisors, divided by the square of the number. [*Submitted by Leo Moser.*]

Q 34. The square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides.

ANSWERS

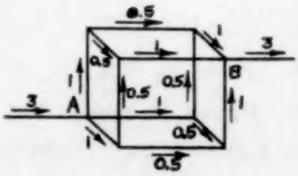
A 28. $(27+8i)/(3+2i^3) = (27-8i^3)/(3-2i) = 9+6i+4i^2 = 5+6i$.

A 29. Since $\log y_n = x \log y_{n-1}$, by iteration we have $\log y_n = x^{n-1} \log x$. Then logarithmic differentiation yields $dy_n/dx = y_n x^{n-2}[1+(n-1)\log x]$.

A 30. Referred to its asymptotes as axes, the equation of a hyperbola is $xy = k$. Let $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ be two points on the hyperbola, and let (u_1, v_1) and (u_2, v_2) be the distances to the asymptotes in the given directions. Then from similar triangles $x_1/x_2 = u_1/u_2$ and $y_1/y_2 = v_1/v_2$. Multiplying, we obtain $u_1v_1/u_2v_2 = 1$, which proves the proposition.

A 31. The locus of the intersection of two perpendicular tangents to the ellipse $b^2x^2 + a^2y^2 = a^2b^2$ is the director circle $x^2 + y^2 = a^2 + b^2$. Hence the center of the moving ellipse which is tangent to two fixed perpendicular lines, considered as axes of reference, is always the distance $(a^2 + b^2)^{\frac{1}{2}}$ from the origin. The closest approach of the center to the axes is at (a,b) and (b,a), so the locus of the center is the minor arc of $x^2 + y^2 = a^2 + b^2$ included between these points. [The length of the arc is $(a^2 + b^2)^{\frac{1}{2}} \arctan (a^2 - b^2)/2ab$.]

A 32. Assume 3 amperes flowing in at A and out at B. The current will divide as shown. Summing up the IR drops along any path gives $1 + 0.5 + 1 = 2.5$ volts. Hence the total resistance is $E/I = 2.5/3$ or 0.833 ohms.



A 33. Dividing a number by a divisor gives a complementary divisor. So if d runs through the divisors of n, n/d will run through these same divisors. Hence the first sum is $\Sigma 1/d$ while the second sum is $\Sigma 1/d^2$, where in both cases d runs through all the divisors of n. Clearly the first is larger except for $n = 1$, when they are equal. (The first sum may be made arbitrarily large by taking $n = m!$, with m sufficiently large. The second sum will always be less than $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$.)

A 34. In the triangle ABC, angle $C = 90°$. With A as center and radius b describe a circle cutting AB at E and BA extended at D. Now BC is tangent to the circle and $AE = b = AD$. Hence $a^2 = (c+b)(c-b)$, so $c^2 = a^2 + b^2$.

OUR CONTRIBUTORS

John W. Green, Associate Professor of Mathematics, University of California, Los Angeles was born in Hearn, Texas in 1914. After graduating from Rice Institute (B.A. '35, M.A. '36) and the University of California (Ph.D. '38) he taught at Harvard and Rochester. From 1943-45 he served as mathematician at Aberdeen Proving Ground and joined the faculty of U.C.L.A. in 1945. An associate editor of the Duke Mathematical Journal, Dr. Green also serves as an Associate Secretary of the American Mathematical Society. During the present year he is spending half time at the Institute of Numerical Analysis, National Bureau of Standards.

Andrew Pikler was born in 1907 in Budapest, Hungary. After studying social sciences at Hungarian, German and French universities he received the Ph.D. degree in 1929 (Pecs). A member of the Econometric Society, Dr. Pikler has published numerous articles in various languages concerning the kinetic theory of money, the economic interpretation of entropy and other related subjects. Dr. Pikler is also active in musical performance and composition, and is interested in the theory of consonance and dissonance in acoustics. He was invited to the United States in 1946 as a "visiting doctor" by the University of Chicago and by Clark University.

Erich Michalup, Professor of Financial Mathematics, University of Caracas, Venezuela, was born in Vienna, Austria in 1902. He is a graduate of the University of Vienna (L.L.D. '30; Actuary and Statistician, '33, Economy and Social Science '36, Ph.D. '38) and is a member of the Swiss and London Institutes of Actuaries, of the American and French mathematical societies. He has also been elected to the Venezuelan Academy of Science and the Inter-American Statistical Institute. Prof. Michalup has published many papers in the fields of mathematics, actuarial science and statistics. His article "On the Summation of Power Series" appeared in the January-February issue.

Alexander W. Boldyreff was born in St. Petersburg, Russia in 1903. He came to America at the age of nineteen and attended Western Reserve University and the University of Michigan (B.S. '26; M.S. '27; Ph.D.'30). He taught mathematics and chemistry at the University of Arizona for about ten years, and also has taught at Illinois, Wittenberg College and the University of New Mexico. Since 1950 he has been on the research staff of the Sandia Corporation, Albuquerque, New Mexico. Dr. Boldyreff is especially interested in "applicable mathematics" and arithmetic number theory. His article, "Decomposition of Rational Fractions into Partial Fractions", appeared in the January-February issue.